University of Alberta

Population genetic structure of the alpine butterfly *Parnassius smintheus*: influence of landscape at small and large spatial scales

by

Nusha Keyghobadi

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Environmental Biology and Ecology

Department of Biological Sciences

Edmonton, Alberta

Fall 2001

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-68957-3

Canada

University of Alberta

Library Release Form

Name of Author: Nusha Keyghobadi

Title of Thesis: Population genetic structure of the alpine butterfly *Parnassius smintheus*: influence of landscape at small and large spatial scales

Degree: Doctor of Philosophy

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

N. Keyghobadi

1900 Sheppard Ave. E. #1205
Willowdale, Ontario
M2J 4T4

August 9, 2001

Birdwatching at the equator

The blue-footed booby
stands on her tropic island
in the Galapagos group
stands all day long
shading her eggs from the sun
also protecting her blue feet
from too much ultraviolet
Sometimes the male booby
flaps his wings and dances
to entertain his mate
pointing his toes upward
so they can discuss blueness
which seems to them very beautiful
Their only real enemy
is the piratical frigate bird
floating on great black wings
above the mile-long island
Sometimes the frigate bird
robs them of their fish
whereupon the booby
is wont to say "Friggit"
and catches some more
When night comes all the boobies
sit down at once as if
God had given them a signal
or else one booby says
to the rest "Let's flop boys"
and they do
The blue booby's own capsule-
comment about evolution:
if God won't do it for you
do it yourself:
stand up
sit down
make love
have some babies
catch fish
dance sometimes
admire your feet
friggit:
what else is there?

- Al Purdy

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Population genetic structure of the alpine butterfly *Parnassius smintheus*: influence of landscape at small and large spatial scales** submitted by Nusha Keyghobadi in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Environmental Biology and Ecology.

Jens Roland

Curtis Strobeck

Bruce Rannala

Felix Sperling

Diane Debinski

Date: Aug 8 /2001

For my parents, Latifeh and Parviz

Abstract

The movement of individuals and their genes has important consequences for ecological and evolutionary dynamics of populations. Biologists are increasingly interested in understanding how movement and gene flow occur in real, heterogeneous landscapes. I investigated the influence of landscape on gene flow and population genetic structure of the alpine butterfly, *Parnassius smintheus* in Banff National Park and Kananaskis Country, Alberta, Canada. Microsatellite DNA markers for *P. smintheus* were developed. As in other butterfly species, development of microsatellites in *P. smintheus* was difficult. This was due to high variability in the regions flanking microsatellites, and possibly to the occurrence of microsatellite sequences within larger, repetitive elements in the genome. At a fine spatial scale (up to 12 km separating sites), I observed a significant negative correlation between genetic distance and movement rates estimated by mark-recapture methods. Furthermore, landscape had a significant influence on population genetic structure; the amount of forested habitat separating populations explained more variation in genetic distance than did the amount of open, meadow habitat. Genetic variation within sites was positively correlated with the connectivity of the site to other local populations of *P. smintheus*. Surprisingly, however, the correlation was greater for a measure of connectivity that did not incorporate differential effects of surrounding meadow and forest habitat than one that did incorporate such effects. This may be due to temporal lags in the response of genetic variation to habitat fragmentation that has occurred in the study area over the past 60 years. Comparison of patterns of genetic structure among three geographic regions revealed that landscape also influences gene flow at a larger spatial scale (up to 58 km separating sites). Among the regions, inferred levels of gene flow increased with increasing connectivity of high-altitude, non-forested habitats. Climatic warming is predicted to cause fragmentation of the alpine meadows occupied by this *P. smintheus* via a rise in the elevational limit of tree-line, restricting meadows to smaller areas of mountaintops. My results suggest that this will lead to reduced movement and gene flow among populations.

Acknowledgments

I am very grateful to my supervisors, Jens Roland and Curtis Strobeck, for the opportunity to work on such a great project and for their intellectual, moral, and financial support throughout. I am glad to have worked with them and to have learned from them both.

Pete Kershaw was a member of my supervisory committee early on and provided helpful ideas at that critical stage. I am also thankful to Bruce Rannala for agreeing to join my committee on late notice, and for his comments on the thesis. Diane Debinski, my external examiner, and Felix Sperling also provided comments that improved the thesis.

This project was generously supported by NSERC, the Killam foundation, Challenge Grants in Biodiversity (Alberta Conservation Association), Alberta Parks and Rec., Banff National Park, and Friends of Banff National Park.

My project took me into both the field and the molecular biology lab, and I am grateful for the range of experiences that it has afforded me. I am also extremely grateful to the many people that have assisted me along my journey. John Coffin provided my early training in the lab and I thank him for his attentiveness and patience. Renee Polziehn taught me about automated sequencing (and even a couple of things about mountaineering). David Paetkau gave me some excellent advice at the beginning of the project, and I am especially thankful to him for his suggestion to try a PCR-based approach to microsatellite isolation. This thesis would have been of an entirely different form if not for his advice. I owe a special debt of gratitude to Corey Davis who generously assisted me and gave me advice on all matters in the lab at all stages of my project. I learned much from him and am grateful for his kindness and good humour. I am also thankful to the other members of the Strobeck lab, who made it a fun place to work and all of whom have helped in some way: Jen Bonneville, Véronique Campbell, Lindsey Carmichael, Dave Coltman, Phyllis Dale, Isabelle Delisle, Jan Jekielek, Chris Kyle, Monique Paradon, Mike Proctor, Maarten Vonhof, Greg Wilson, and Keri Zittlau.

I am very lucky to have worked with Sherri Fownes. She was an excellent field assistant and a good friend and I thank her for her warmth and kindness. I am also thankful for the many important things she taught me about *Parnassius* biology. I thank Steve Matter for running the Virtual Migration Model - and explaining how it works - and for discussions

on the third chapter of my thesis. I am also thankful to all the other members of the Roland lab, past and present, with whom I have enjoyed interacting very much. They are a fun and generous bunch: Corina Brdar, Helena Bylund, Barry Cooke, Ed Mondor, April Mitchell, Lorne Rothman, Kris Sabourin, Chris Schmidt, Kirsty Ward and Brian Vanhezewijk.

The following people assisted with sample collection and mark-recapture studies and I would have been at a loss without their help: Norine Ambrose, John Brzustowski, Mark Caldwell, Sue Cotterill, Alison Crysler, Nicholas DeCarlo, Marian Forrester, Sherri Fownes, Fiona Johnson, Kris Sabourin, Chris Schmidt, Lynette Scott, Kirsty Ward, and Arliss Winship. I am thankful to staff of the Kananaskis Field Stations for their logistical support, especially Judy Buchanan-Mappin and Grace LeBel. I am also grateful to the staff of Alpine Helicopters in Canmore for their friendly service and for providing us with both fruitful collecting trips and exciting adventures. I would like to thank Cliff White from Banff National Park for his support.

I thank John Brzustowski for writing the program to perform the EM algorithm, and for the other programs that he had written previously and of which I made extensive use.

I am thankful to the Entomology grad students and post-docs with whom I shared an office during the writing of this thesis. Their friendliness and light-heartedness made the writing much less painful than it might have been otherwise.

I would like to thank my parents, Latifeh and Parviz, my brother, Kiarash, and my aunt, Fereshteh, and all the members of my family for their support and encouragement. I have missed them dearly. I thank my friends and my soccer buddies for taking my mind off my work every once in a while and for providing me with a community here in Edmonton.

I thank Graeme Taylor for many, many things - for encouraging me to pursue a doctoral degree, for taking care of me along the way, for advice on all matters, academic or otherwise, and for his companionship. Graeme is one of the most passionate biologists I have ever met and his love for the subject has been an inspiration to me. Thanks for taking me outside the dotted lines ...

Table of Contents

1. General Introduction 1
- Subdivided populations and individual movement 1
- Subdivided populations in real landscapes 2
- Overview of thesis 3
- Why study *Parnassius smintheus* ? 4
- Taxonomic status and distribution of *Parnassius smintheus* 5
- Natural history of *Parnassius smintheus* 6
- Microsatellite loci as markers for analysis of population genetic structure 8
- Literature cited 10

2. Influence of landscape on the fine-scale population genetic structure of the alpine butterfly, *Parnassius smintheus* 14
- Introduction 14
- Materials and methods 17
  - *Isolation of microsatellites* 17
  - *Sample collection and DNA isolation* 19
  - *PCR amplification and analysis of microsatellites* 19
  - *Landscape variables* 20
  - *Estimating butterfly movements* 20
  - *Statistical analysis* 21
- Results 22
  - *Isolation of microsatellites* 22
  - *Hardy-Weinberg equilibrium and evidence for null alleles* 23
  - *Maximum likelihood estimation of allele frequencies* 24
  - *Mantel tests and the effects of landscape structure on genetic structure* .. 25
  - *Analysis restricted to Jumpingpound Ridge and Cox Hill* 26
- Discussion 27
  - *Microsatellites in butterflies* 27
  - *Dispersal rates and genetic distances* 28
  - *Genetic structure at a small spatial scale* 29
  - *Landscape structure and population genetic structure* 30
  - *Implications for long-term persistence* 31
- Literature cited 32
- Appendix 2-A 45

3. Habitat fragmentation and genetic variation: effects of climatically induced rising tree-line on an alpine butterfly 47
- Introduction 47
- Materials and methods 49
  - *Study area* 49
  - *Mark-recapture data and estimation of population size* 50
  - *Measurement of landscape variables* 51

Measures of population size 51
Index of patch connectivity 51
Sample collection, DNA extraction, and microsatellite genotyping 53
Statistical analysis 54
Results 55
Effects of intervening landscape on genetic differentiation among sites ... 55
Levels of genetic variation within sites 56
Genetic variation and population size 56
Genetic variation and patch connectivity 57
Correlation of heterozygosity with patch areas and connectivities from 1952 58
Discussion 59
Hypothesis 1: Genetic variation and population size 59
Hypothesis 2: Genetic variation and patch connectivity 61
Hypothesis 3: Importance of intervening landscape as a component of patch connectivity 61
Sampling units 63
Selection and mutation 64
Immigration from outside the study area 65
Long-term implications of habitat fragmentation for P. smintheus 65
Literature cited 67

4. Large-scale patterns of population differentiation and landscape connectivity in the alpine butterfly, *Parnassius smintheus* 80
Introduction 80
Materials and methods 84
Study area 84
Measurement of landscape variables 84
Sample collection and microsatellite genotyping 85
Data analysis 86
Comparison to genetic structure at a smaller spatial scale 88
Results 88
Landscape connectivity 88
Hardy-Weinberg equilibrium 89
Microsatellite variability 89
Genetic structure and F „ estimates 90
Isolation by distance 90
Differences in within-population genetic variation among regions 90
Discussion 91
Gene flow versus genetic drift 91
Lack of isolation by distance in the foothills landscape 92
Comparison of Banff and West Kananaskis regions 94
Historical changes in landscape structure 96
Low estimated values of F „ and high variability of microsatellites 97
Conclusion 98
Literature cited 100

5. General Conclusion 110
Landscape and population genetic structure 110
Microsatellites for analysis of population structure in butterflies 111
Estimation of movement rates at a large spatial scale 112
Alternative methods of genetic data analysis 114
Non-neutral variation 115
Habitat of *Parnassius smintheus* 115
Butterflies and climate change 116
Literature cited 118

6. Microsatellite loci in a butterfly: flanking sequence variability and other impediments to marker development (Addendum) 121
Introduction 121
Results 123
Isolation of novel microsatellite markers 123
Variability, Hardy-Weinberg equilibrium, and null alleles 124
P. smintheus *microsatellites: sequence, length, and utility for genotyping individuals* 125
Sequencing of microsatellites and flanking regions 127
Discussion 128
Amplification of microsatellite loci 128
Homozygote excess and null alleles 129
Variability of microsatellite flanking regions 131
Experimental procedures 134
Microsatellite isolation 134
Sample collection and DNA extraction 136
Microsatellite amplification and analysis 136
Sequencing of microsatellites and flanking regions 137
Literature cited 138

List of Tables

Table 2-1. Primer sequences for four *P. smintheus* microsatellite loci..... 36
Table 2-2. Genetic variation within each sampling site for fine-scale study..... 37
Table 2-3. Results of simple and partial Mantel tests..... 38
Table 3-1. Genetic variation, population size, and connectivity of study sites..... 72
Table 4-1. Genetic variation within sample sites for large-scale study..... 103
Table 4-2. Landscape characteristics and population genetic structure..... 104
Table 6-1. Novel microsatellite loci for *P. smintheus* 141
Table 6-2. Amplification results for microsatellites isolated from *P. smintheus* 142

List of Figures

Figure 2-1. Locations of sampling sites..... 39
Figure 2-2. Evidence for null alleles at locus Ps50..... 40
Figure 2-3. Genetic distance versus proportion of marks moving between sites..... 41
Figure 2-4. Genetic distance versus total geographic distance..... 42
Figure 2-5. Genetic distance versus total elevation difference..... 43
Figure 2-6. Genetic distance versus residuals from regressions..... 44
Figure 3-1. Study area..... 73
Figure 3-2. Genetic variation versus measures of population size..... 74
Figure 3-3. Genetic variation versus population size for combined sites..... 75
Figure 3-4. Genetic variation versus patch connectivity (1993)..... 76
Figure 3-5. Genetic variation versus connectivity based on total distance (1993)..... 77
Figure 3-6. Aerial photographs of Jumpingpound Ridge..... 78
Figure 3-7. Genetic variation versus patch connectivity (1952)..... 79
Figure 4-1. Locations of sample sites..... 105
Figure 4-2. Isolation by distance in West Kananaskis..... 106
Figure 4-3. Isolation by distance in Banff..... 107
Figure 4-4. Isolation by distance in East Kananaskis..... 108
Figure 4-5. Isolation by distance in small-scale study area..... 109
Figure 5-1. Estimation of movement rates at a large spatial scale..... 120
Figure 6-1. Aligned sequences of alleles for four microsatellite loci..... 143

Chapter 1:
General Introduction

Subdivided populations and individual movement

Individuals in natural populations are rarely distributed uniformly in space, and at some scale most populations can be described as "patchy" (Andrewartha & Birch 1954). If interactions among individuals consistently occur more frequently within patches than between patches, then the population is considered to be subdivided into independent or semi-independent demographic units. Population subdivision has important consequences for ecological and evolutionary dynamics which have long been recognized (Nicholson 1933, Andrewartha & Birch 1954, Wright 1931). Population subdivision may affect single-species population dynamics (Levins 1970, Hanski & Gilpin 1991, Hastings 1991), interspecific competitive interactions (Caswell & Cohen 1991, Bengtsson 1991), predator-prey dynamics (Nicholson 1933, Hassell *et al.* 1991) and other trophic interactions (Polis *et al.* 1997), as well as effective population size, genetic variation, and adaptation (Wright 1978, Varvio *et al.* 1986, Gilpin 1991). Population subdivision and spatial structure are also recurring themes in conservation biology because anthropogenic fragmentation of natural habitats is considered one of the most prevalent causes of species decline and extinction (Forman 1995).

A critical process affecting the dynamics of subdivided populations is movement of individuals among local demographic units. For example, a certain amount of movement among habitat patches may allow populations to persist in stochastic environments where they would otherwise be highly susceptible to extinction (Stacey *et al.* 1997). Similarly, gene flow, which is mediated by individual movements, is a fundamental microevolutionary force that can determine the degree of genetic differentiation among local demographic units and affect the course of adaptation within species (Slatkin 1985).

Subdivided populations in real landscapes

Early models of subdivided populations such as Levins' (1970) model of metapopulations or Wright's (1931, 1978) island model of migration do not consider the spatial configuration of local demographic units and do not incorporate any spatial component to movements of individuals or genes. Even though local demographic units are considered to be discrete, they are all assumed to be equally connected to each other by individual movements. More sophisticated models that incorporate spatial relations among local demographic units have since been developed. In ecology, spatially explicit models of metapopulation dynamics have been described and tested (Hanski 1994). These models consider movement between habitat patches to be a function of patch sizes and of the distances separating patches. In population genetics, models of stepping-stone migration (Kimura & Weiss 1964) have added a spatial component to the analysis of genetic variation within and among local populations. These models restrict movement and therefore, gene flow, to adjacent demographic units in a one- or two-dimensional array.

Although many theoretical and empirical studies of subdivided populations have incorporated the spatial relations among demographic units, few have considered other features of landscape structure. The sizes of habitat patches or demographic units and the distances between them are considered, but not other variation in the landscape, such as the quality of habitat patches or the type of terrain separating patches (Wiens 1997). Thus, habitat patches are often assumed to be surrounded by a homogeneous matrix. However, the qualities of the landscape separating local demographic units can affect movement of individuals and genes, thereby exerting an important influence over the ecological and evolutionary dynamics of subdivided populations. There is growing interest in describing and understanding the detailed effects of landscape on movement and the consequences for subdivided populations (Wiens 1997, Sork *et al.* 1999, Debinski *et al.* 2001). An understanding of such effects is necessary for building more realistic population models,

for more precise testing of theory, and for predicting the impacts of anthropogenic changes to the landscape on populations.

Overview of thesis

This thesis is an empirical analysis of the influence of landscape on population genetic structure and gene flow in the alpine butterfly *Parnassius smintheus* Doubleday (1847). Here, I test the hypothesis that different landscape elements differentially affect rates of gene flow and dispersal, and that the distribution of those elements are important determinants of the levels and distribution of genetic variation. Two particular strengths of this study are the integration of population genetic data with detailed data on individual movement from a mark-release-recapture (MRR) study (Roland *et al.* 2000) and the analysis of population genetic structure at two spatial scales.

In Chapter 2, I test the hypothesis that the qualities of the landscape separating local populations affect the degree of genetic differentiation among those populations. I examine genetic distances between pairs of sample sites in a fine-scale study area in relation to the qualities of the intervening landscape. Another aim of this chapter is to integrate the genetic data with MRR data collected from the same set of sample sites (Roland *et al.* 2000). Thus, I ascertain whether rates of movement between sites, as determined by MRR, are related to the degree of genetic differentiation between sites. Furthermore, I determine the extent to which the effects of landscape on individual movements observed using MRR (Roland *et al.* 2000) are translated into effects of landscape on gene flow and population genetic structure.

Rates of gene flow among demographic units in a subdivided population affect not only the degree of genetic differentiation among sites, but also levels of genetic variation within sites. Having examined the relationships between landscape and genetic differentiation among sample sites in the fine-scale study area (Chapter 2), in Chapter 3 I aim to determine whether genetic variation within sample sites is affected by the degree of

isolation of the site and the size of the local population. In particular, the correlations between within-site heterozygosity and measures of site isolation that both do and do not account for differential effects of surrounding meadow and forest habitat on movement are examined.

In Chapter 4, I examine whether the effects of different landscape elements on dispersal (Roland *et al.* 2000) and genetic structure (Chapters 2 & 3) that are observed at the small spatial scale of the MRR study (< 12 km separating sample sites) are also detectable at a larger spatial scale (up to 58 km separating sites) at which it is not feasible to conduct MRR studies. Here, I compare the degree of genetic differentiation among sampled populations, levels of genetic variation within populations, and patterns of isolation by distance among three geographic regions differing in the coverage and connectivity of high-altitude, non-forested habitat.

Why study *Parnassius smintheus*?

Butterflies have served as model systems for the study of subdivided populations. Butterfly population biologists have devoted considerable effort to studying population spatial structure and dispersal within and among patches (e.g., Ehrlich 1961, Ehrlich 1965, Brussard & Ehrlich 1970, Ehrlich & Gilbert 1973, Gilbert & Singer 1973, Watt *et al.* 1977, Shapiro 1979, Harrison *et al.* 1988, Thomas & Harrison 1992, Hill *et al.* 1996, Lewis *et al.* 1997). This is due in part to the fairly high habitat specificity of many butterfly species that makes the patchy distribution of populations in the landscape quite obvious. Also, extensive data on natural history and distributions of butterflies are available, thanks to the work of a large body of amateur collectors. Some of the earliest studies of structured populations that considered both ecology and genetics were on butterflies (see Ford 1964 for review). The study of subdivided populations in butterflies has a long history and provides a large body of work against which one can compare results.

P. smintheus occupies alpine and subalpine meadows that are patchily distributed in the landscape, being separated to varying degrees by montane forests. Within a given meadow, individuals tend to be localized to "hot spots" and there are certain meadows or portions of meadows in which *P. smintheus* is rarely, if ever, seen (pers. obs.). Thus, *P. smintheus* is an ideal subject for a study of the effects of different habitat types on dispersal and gene flow because the patchy structure of populations is generally obvious and movements between local populations can occur through both meadow and forest habitat.

Although *P. smintheus* is both widespread and locally abundant in North America, its congeners *Parnassius apollo* and *Parnassius mnemosyne* are listed as threatened or endangered in Europe (Meglécz *et al.* 1997, Witkowski *et al.* 1997). The decline of both species is partly attributed to a reduction in the area of open, meadow habitat and an increase in forest cover in the landscape (Witkowski *et al.* 1997, Kuras *et al.* 2000). In North American alpine areas, climatic warming may be causing a rise in the elevational limit of tree growth and a similar reduction in the amount of meadow habitat above tree-line (Kearney 1982, Taylor 1995, Woodward *et al.* 1995). If the current state of European Parnassians is any indication, *P. smintheus* may be vulnerable to such a landscape change. Thus, there is also a practical motivation to studying differential effects of forested and non-forested habitat on population structure in *P. smintheus,* to better understand the implications of climatically induced changes in landscape for this species. However, as populations of *P. smintheus* are still large and widespread, the study of this species is not encumbered by the difficulties that accompany research on rare or endangered species.

Taxonomic status and distribution of *Parnassius smintheus*

The genus *Parnassius* Latreille (1804) has a Holarctic distribution with the majority of species found in the Palearctic region. The center of diversity for the genus is in the mountainous areas of central and eastern Asia. All North American species are found in the western part of the continent, mostly in mountainous areas. Until recently, *P. smintheus*

in Alberta was considered a sub-species of *Parnassius phoebus* Fabricius (1793) which has a wide range covering much of northern Eurasia and extending into North America. However, based on variation in the structure of the egg micropyle, the opening through which the egg becomes fertilized, Shepard & Manley (1998) elevated *P. phoebus smintheus* to *P. smintheus*. In addition, they placed a number of other former North American subspecies of *P. phoebus* within *P. smintheus*. The range of *P. smintheus* extends along the Rocky Mountain cordillera from the southern Yukon Territory to northern New Mexico with disjunct populations in the Coast Mountains and North Cascade Mountains of British Columbia and Washington, the Olympic Peninsula, Vancouver Island, the Black Hills, and the Siskiyou Mountains (Shepard & Manley, 1998). *P. phoebus* is still considered to occur in North America, in Alaska and the Yukon Territory. Shepard & Manley (1998) list nine subspecies of *P. smintheus* but state that their status is unclear. However, all populations in Alberta are generally considered to belong to *P. smintheus smintheus* (Shepard & Shepard 1975) which Shepard & Manley (1998) suggest is a valid subspecies.

In Alberta, *P. smintheus* is found throughout the Rocky Mountains (Sperling & Kondla 1991). It mostly inhabits alpine and subalpine meadows containing its host plant, stonecrop (*Sedum lanceolatum*). In our study area in Kananaskis and eastern Banff National Park populations of *P. smintheus* are found just above tree-line. Here, they are most commonly found on steep slopes (> 30°) with well-drained, sandy soils at elevations of 1900 m to 2200 m (Fownes 1999). At more southerly locations in the province, populations are also found at lower elevations (Sperling & Kondla 1991).

Natural history of *Parnassius smintheus*

P. smintheus passes through one generation per year (Scott 1973, Sperling & Kondla 1991). Overwintering occurs in the egg stage (Fownes 1999). Larvae eclose in the spring and pass through five instars (Fownes 1999) before they burrow into pebbly soil where

they pupate (Scott 1986). In Alberta, adults may be found flying from June to October (Sperling & Kondla 1991), although in our study area the flight season is usually from July to September. Scott (1973) reported that adult males survive for an average of six days and the results of Roland *et al.* (2000) suggest that adults live for an average of approximately ten days. In both studies, some individual adults were observed to survive for over three weeks.

P. smintheus larvae have been reported to feed on a number of *Sedum* species (Crassulaceae): *S. lanceolatum, S. obtustatum*, and *S. rosea* (Scott 1973). However, larvae have only been found feeding on *S. lanceolatum* in Colorado (Scott 1973) and Alberta (Bird *et al.* 1995). In Kananaskis, Alberta, Fownes (1999) only found larvae feeding on *S. lanceolatum*, even though *S. rosea* also occurred in those meadows. Larvae sequester a protective cyanoglucoside, sarmentosin, from the host plant (Nishida & Rothschild 1995). This chemical has been found in tissues of all life stages and is thought to provide protection against vertebrate predators (Nishida & Rothschild 1995). Adults feed on nectar from flowers of a variety of plant species, including *S. lanceolatum* (Fownes 1999). Most other nectar plants are members of the Rosaceae and Asteraceae (Fownes 1999).

Female *P. smintheus* display the interesting behaviour of not ovipositing on the host plant, a behaviour shared with other members of the genus such as *P. apollo, P. clodius,* and *P. mnemosyne* (Scott 1973, Fownes 1999). Females will lay their eggs singly and randomly on grasses, other plants, stones, sticks, and soil (Scott 1973, Fownes 1999). Females do not even attempt to oviposit near the host plant, although studies of caged females did suggest that the presence of the host plant promotes oviposition (Fownes 1999). Therefore, first-instar larvae must search for the host plant upon eclosion. Fownes (1999) found that first-instar larvae are mobile and tend to orient towards the host plant.

Adult males are much more conspicuous than females. Males patrol in search of females, flying approximately 30 cm to 1 m above the ground (Scott 1973; Sperling &

Kondla 1991), while females tend to spend less time flying and more of their time feeding, sitting, or ovipositing (Scott 1973). Patrolling males search for females using both visual and olfactory cues (Scott 1973). Copulation and can last for several hours (Scott 1973). During the lengthy copulation, the male deposits a hard, waxy structure called a 'sphragis' over the female's abdomen. This structure prevents the female from mating again and may also plug the female's pheromone glands (Scott 1973).

Adult *P. smintheus* display low vagility relative to many other butterfly species. Scott (1973) found that most adults moved less than 100 m and Roland *et al.* (2000) found an average movement during a season of approximately 150 m. However, long-distance movements may occur and Sperling (1993) reported two independent sightings of adults flying almost 15 km from the nearest patch of their normal habitat. Flight is strongly affected by local weather conditions. Adults do not fly during cloudy periods and winds stronger than approximately 7 m/sec reduce flight activity considerably (Scott, 1973).

Microsatellite loci as markers for analysis of population genetic structure

A microsatellite DNA locus is a tandemly repeated string of a short sequence motif (1-6 bp in length), occurring at a particular location in the genome. Variation among alleles usually involves differences in the number of times the motif is repeated and therefore, in the length of the segment of DNA. Hence, microsatellites belong to a class of markers known as Variable Number of Tandem Repeat (VNTR) markers. Microsatellites were first described in the late 1980s (Litt & Luty 1989, Tautz 1989, Weber & May 1989). Over the past decade they have become the genetic marker of choice for a variety of applications including linkage mapping, identification of individuals or 'DNA fingerprinting', and studies of population biology.

Currently, a variety of different genetic markers are available for assaying heritable variation within and among populations. Each marker has a particular suite of characteristics with respect to levels of variation, mode of inheritance, selective neutrality,

and ease of development. The choice of the optimal marker for any particular study will depend on trade-offs among these characters in relation to the questions being addressed and the spatial and temporal scale of analysis. Of particular importance is the rate of change of variation at a particular marker, which should correspond to the time scale of the ecological or evolutionary process being investigated (Sunnucks 2000).

Allele frequencies at microsatellite loci are largely altered by genetic drift, gene flow, and a high rate of mutation (Sunnucks 2000). Therefore, these markers are appropriate for analysis of population structure and gene flow. Also, because microsatellites are highly variable, they are particularly useful for analysis of population structure and population-level processes at relatively fine spatial and temporal scales, even in species with limited variation at other genetic markers (Hughes & Queller 1993). Other desirable features of microsatellites are that they are generally considered to be selectively neutral (Jarne & Lagoda 1996), they can be assayed by the polymerase chain reaction (PCR), and they are single-locus, co-dominant markers. Based on this suite of characters, microsatellite loci were considered the most appropriate markers for this study. It became clear through the course of this study, from my results and those of other butterfly researchers (Nève & Méglecz 2000), that the development of microsatellite DNA markers in butterfly species is particularly challenging. In Chapter 2, I describe the development of four microsatellite loci for *P. smintheus*. In Chapter 6 (Addendum), I describe the development of four additional loci and also investigate potential reasons for the difficulty of development of microsatellite markers in this and possibly other butterfly species. Despite the difficulties associated with microsatellite development in this group of species, the pursuit seemed worthwhile as microsatellites held the potential for detecting genetic structure at a very fine spatial scale, allowing the unique integration of genetic data with ecological data from MRR.

Literature Cited

Andrewartha H. G. and Birch L. C. (1954) The Distribution and Abundance of Animals. University of Chicago Press, Chicago.

Bengtsson J. (1991) Interspecific competition in metapopulations. *Biological Journal of the Linnean Society* **42**: 219-238.

Bird C. D., Hilchie G. J., Kondla N. G., Pike E. M. and Sperling F. A. H. (1995) Alberta butterflies. Provincial Museum of Alberta, Edmonton.

Brussard P. F. and Ehrlich P. R. (1970) The population structure of *Erebia epipsodea*. *Ecology* **51**: 119-129.

Caswell H. and Cohen J. E. (1991) Disturbance, interpecific interaction, and diversity in metapopulations. *Biological Journal of the Linnean Society* **42**: 193-218.

Debinski D. M., Ray C., and Saveraid E. H. (2001) Species diversity and the scale of the landscape mosaic: do scales of movement and patch size affect diversity? *Biological Conservation* **98**: 179-190.

Ehrlich P. R. (1961) Intrinsic barriers to dispersal in the checkerspot butterfly, *Euphydryas editha*. *Science* **134**: 108-109.

Ehrlich P. R. (1965) The population biology of the butterfly, *Euphydryas editha*.. II. The structure of the Jasper Ridge colony. *Evolution* **19**: 327-336.

Ehrlich P. R. and Gilbert L. E. (1973) Population structure and dynamics of the tropical butterfly *Heliconius ethilla*. *Biotropica* **5**: 69-82.

Ford E. B. (1964) Ecological Genetics. Methuen & Co. Ltd., London.

Forman R. T. T. (1995) Land Mosaics: the Ecology of Landscapes and Regions. Cambridge University Press, Cambridge.

Fownes S. L. (1999) Meadow suitability and quality for the alpine butterfly, *Parnassius smintheus*, in the east slopes of the Canadian Rockies. M.Sc. Thesis. University of Alberta.

Gilbert L. E. and Singer M. C. (1973) Dispersal and gene flow in a butterfly species. *American Naturalist* **107**: 58-72.

Gilpin M. (1991) The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society* **42**: 165-175.

Hanski I. (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology* **63**: 151-162.

Hanski I. and Gilpin M. (1991) Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society* **42**: 3-16.

Harrison S., Murphy D. D. and Ehrlich P. R. (1988) Distribution of the Bay Checkerspot Butterfly, *Euphydryas editha bayensis*: evidence for a metapopulation model. *American Naturalist* **132**: 360-382.

Hassell M. P., May R. M., Pacala S. W. and Chesson P. L. (1991) The persistence of host-parasitoid associations in patchy environments I. A general criterion. *American Naturalist* **138**: 568-583.

Hastings A. (1991) Structured models of metapopulation dynamics: concepts, models, and observations. *Biological Journal of the Linnean Society* **42**: 57-72.

Hill J. K., Thomas C. D. and Lewis O. T. (1996) Effects of habitat patch size and isolation on dispersal by *Hesperia comma* butterflies: implications for metapopulation structure. *Journal of Animal Ecology* **65**: 725-735.

Hughes C. R. and Queller D. C. (1993) Detection of highly polymorphic microsatellite loci in a species with little allozyme polymorphism. *Molecular Ecology* **2**: 131-137.

Jarne P. and Lagoda P. J. L. (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**: 424-429.

Kearney M. S. (1982) Recent seedling establishment at timberline in Jasper National Park, Alta. *Canadian Journal of Botany* **60**: 2283-2287.

Kimura M. and Weiss G. H. (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**: 561-576.

Kuras T., Benes J., Celechovsky A., Vrabec V. and Konvicka M. (2000) *Parnassius mnemosyne* (Lepidoptera: Papilionidae) in North Moravia: review of present and past distribution, proposal for conservation. *Klapalekiana* **36**: 93-112.

Levins R. (1970) *Extinction. in* Gerstenhaber, M. (ed.), Lectures on Mathematics in the Life Sciences. Vol 2. American Mathematical Society, Providence, R. I.

Lewis O.T., Thomas C.D., Hill J.K., Brookes M. I., Crane T. P. R., Graneau Y. A., Mallet J. L. B. and Rose O. C. (1997) Three ways of assessing metapopulation structure in the butterfly *Plebejus argus*. *Ecological Entomology* **22**: 283-293.

Litt M. and Luty J. A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**: 397-401.

Meglécz E., Pecsenye K., Peregovits L. and Varga Z. (1997) Effects of population size and habitat fragmentation on the genetic variability of *Parnassius mnemosyne* populations in NE Hungary. *Acta Zoologica Academiae Scientiarum Hungaricae* **43**: 183-190.

Nève G. and Méglecz E. (2000) Microsatellite frequencies in different taxa. *Trends in Ecology and Evolution* **15**: 376-377.

Nicholson A. J. (1933) The balance of animal populations. *Journal of Animal Ecology* **2**: 132-178.

Nishida R. and Rothschild M. (1995) A cyanoglucoside stored by a *Sedum*-feeding Apollo butterfly, *Parnassius phoebus*. *Experientia* **51**: 267-269.

Polis G. A, Anderson W. B. and Holt R. D. (1997) Toward an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs. *Annual Review of Ecology and Systematics* **28**: 289-316.

Roland J., Keyghobadi N. and Fownes S. (2000) Alpine *Parnassius* butterfly dispersal: effects of landscape and population size. *Ecology*.**81**: 1642-1653.

Scott J. A. (1973) Population biology and adult behavior of the circumpolar butterfly *Parnassius phoebus* F. (Papilionidae). *Entomologia Scandinavia* **4**:161-168.

Scott J. A. (1986) The Butterflies of North America: A Natural History and Field Guide. Stanford University Press, Stanford.

Shapiro A. M. (1979) Weather and the lability of breeding populations of the checkered white butterfly, *Pieris protodice* Boisduval and LeConte. *Journal of Research on the Lepidoptera* **17**: 1-23.

Shepard J. H. and Shepard S. S. (1975) Subfamily Parnassiinae. pp. 403-409. *In* W. H. Howe (ed.) The Butterflies of North America. Doubleday & Co., New York.

Shepard J. H. and Manley T. R. (1998) A species revision of the *Parnassius phoebus* complex in North America (Lepidoptera: Papilionidae). pp. 717-726. *In* T. C. Emmel (ed.) Systematics of Western North American Butterflies. Mariposa Press, Gainesville.

Slatkin M. (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**:393-430.

Sork V. L., Nason J., Campbell D. R. and Fernandez J. F. (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution* **14**: 219-224.

Sperling, F. A. H. (1993) Twenty-seven years of butterfly observations at Fish Butte, Alberta. *Blue Jay* **51**: 132-137.

Sperling F. A. H.and Kondla N. G. (1991) Alberta Swallowtails and Parnassians: Natural history, keys, and distribution. *Blue Jay* **49**:183-192.

Stacey P. B., Johnson V. A. and Taper M. L. (1997) Migration within metapopulations: the impact upon local population dynamics. pp. 267-291 *In* Hanski I. A. and Gilpin M. E. (eds.) Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, San Diego.

Sunnucks P. (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution* **15**: 199-203.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463-6471.

Taylor A. H. (1995) Forest expansion and climate change in the mountain hemlock (*Tsuga mertensiana*) zone, Lassen Volcanic National Park, California, USA. *Arctic and Alpine Research* **27**: 207-216.

Thomas C. D. and Harrison S. (1992) Spatial dynamics of a patchily distributed butterfly species. *Journal of Animal Ecology* **61**: 437-446.

Varvio S., Chakraborty R. and Nei M. (1986) Genetic variation in subdivided populations and conservation genetics. *Heredity* **57**: 189-198.

Watt W. B., Chew F. S., Snyder L. R. G., Watt A. G. and Rothschild D. E. (1977) Population structure of Pierid butterflies I. Numbers and movements of some montane *Cloias* species. *Oecologia* **27**: 1-22.

Weber J. L. and May P. E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.

Wiens J. A. (1997) Metapopulation dynamics and landscape ecology. pp. 43-62 *In* Hanski I. A. and Gilpin M. E. (eds.) Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, San Diego.

Witkowski Z., Adamski P., Kosior A. and Plonka P. (1997) Extinction and reintroduction of *Parnassius apollo* in the Pieniny National Park (Polish Carpathians). *Biologia (Bratislava)* **52**: 199-208.

Woodward A., Schreiner E. G. and Silsbee D. G. (1995) Climate, geography, and tree establishment in subalpine meadows of the Olympic Mountains, Washington, U.S.A. *Arctic and Alpine Research* **27**: 217-225.

Wright S. (1931) Evolution in Mendelian populations. *Genetics* **16**: 97-159.

Wright S. (1978) Evolution and the Genetics of Populations. Vol. 4: Variability Within and Among Natural Populations. University of Chicago Press, Chicago.

Chapter 2:

Influence of landscape on the fine-scale population genetic structure of the alpine butterfly, *Parnassius smintheus* *

Introduction

Dispersal rates of animals may differ through different habitat types and the nature of the landscape separating populations can have a strong influence on the degree of their genetic differentiation (Avise & Felley 1979; Pounds & Jackson 1981; Caccone & Sbordoni 1987; King 1987; Arter 1990; Kane *et al.* 1992; Preziosi & Fairbairn 1992; Napolitano & Descimon 1994; Britten *et al.* 1995; Johnson & Black 1995; Johannesen *et al.* 1996; Nève *et al.* 1996; Meglécz *et al.* 1997; VanDongen *et al.* 1998). The potential influence of landscape on population differentiation has long been recognized (Fisher & Ford 1947), its ultimate expression being allopatric speciation via geographic isolation (Mayr 1942). Molecular tools for population genetic analysis allow us to readily quantify the effects of landscape structure on geographic patterns of genetic variation. The determination of such effects can lead to more precise testing of population genetic theories, and allows us to predict the impacts of anthropogenic changes to the landscape on the genetic health of species.

More recently, effects of landscape structure on patterns of genetic variation have been demonstrated in a number of ways. For example, populations separated by a putative physical barrier are found to be more genetically dissimilar than populations not separated by such a barrier (e.g., Avise & Felley 1979; Kane *et al.* 1992; Preziosi & Fairbairn 1992; Napolitano & Descimon 1994; Johannesen *et al.* 1996; Nève *et al.* 1996; Meglécz *et al.* 1997). In most studies of this type, identifying a relationship between landscape features

---

* A version of this chapter has been published. Keyghobadi, Roland, and Strobeck. (1999) *Molecular Ecology* 8: 1481-1495.

and the distribution of genetic variation was not the original goal, but was demonstrated after the fact. Although such results are convincing, the post-hoc approach has its drawbacks. First, confounding factors, geographic distance in particular, are often not controlled. Second, there is a tendency to notice the obvious and discrete structures that may act as barriers to gene flow, such as valleys, ridges, or rivers. Subtle differences in habitat type that may affect gene flow are less often considered (but see Meglécz *et al.* 1997).

Other studies have compared patterns of genetic variation between regions with different landscape structure (Caccone & Sbordoni 1987; Britten *et al.* 1995; Johnson & Black 1995; VanDongen *et al.* 1998). For example, Britten *et al.* (1995) ascribed the low levels of gene flow and lack of isolation-by-distance among populations of the butterfly *Euphydryas editha* in the Great Basin, compared with those in the Rocky Mountains, to the patchy distribution of habitat in the Great Basin area. They suggested that in the Great Basin the species' high-elevation, boreal habitat is highly fragmented by unsuitable habitat at lower elevations, thus limiting the ability of adults to disperse between populations. In most of these studies, assessing the relationship between landscape structure and genetic structure was one of the primary goals.

King (1987) and Arter (1990) took a unique approach to demonstrating how geographic patterns of genetic variation are influenced by landscape features. Both examined the associations between a matrix of pairwise genetic distances and several matrices of pairwise "geographical" distances, using the Mantel test of matrix correspondence (Mantel 1967) which is analogous to a test of simple correlation. Each matrix of "geographical" distance represents a different hypothesis regarding a particular route through the landscape along which gene flow might occur. For example, in her study of a beetle species, King (1987) used a binary matrix that assigned a value of zero to pairs of populations separated by a river and a value of one to pairs of populations not so separated; this matrix represented the hypothesis that rivers acted as barriers to gene flow.

In his study of an alpine snail, Arter (1990) used a matrix of pairwise distances based on the paths that would have to be taken if all movements were restricted to following the drainage system. The degree of association between the genetic distance matrix and a particular matrix of "geographical" distances determines the support for the hypothesis represented by that geographical distance matrix.

In this study, we investigated how genetic distances between populations and sub-populations of an alpine butterfly, *Parnassius smintheus* Doubleday (1847), are affected by the nature of the intervening landscape. We investigated the associations between pairwise genetic distances and a number of pairwise distances based on landscape variables using simple Mantel and partial Mantel tests (Smouse *et al.* 1986). Our approach was similar to that of King (1987) and Arter (1990); by using partial Mantel tests, which are analogous to calculating partial correlation coefficients, we controlled for associations between our measures of landscape structure.

P. smintheus inhabits alpine and sub-alpine meadows that support its larval host plant *Sedum lanceolatum* (Scott 1973; Sperling & Kondla 1991). Its occurrence in meadows that have a patchy distribution in the mountain landscape makes it an ideal subject for studies of population subdivision, dispersal, and gene flow. *P. smintheus* is univoltine and in our study area in south-western Alberta, Canada, adults are in flight from July until September and nectar-feed on a variety of plants. Males patrol in search of females (Scott 1973; Sperling & Kondla 1991), and there is no evidence of hill-topping in this species.

Using mark-release-recapture, Roland *et al.* (2000) studied the effects of landscape structure and population size on dispersal of adult *P. smintheus*. Rates of movement between habitat patches were modelled statistically as a function of the distance between the two patches that is through forest, the distance through open meadow, the total change in elevation between them (both up and down), the areas of both patches, and the population sizes in both patches. The distance through forest, the distance through meadow, and the elevation change had significant effects on movement rates. The distance through forest

had the strongest effect and rates of movement through forest were greatly reduced compared to rates of movement through meadow habitat. We examined whether the effects of landscape on dispersal are reflected in the relationship between landscape structure and population genetic structure.

We describe the genetic relationships among the same sites used in the mark-recapture study of Roland *et al.* (2000). The scale of this study is small, with less than 12 km separating the most distant populations (Fig. 2-1). The detection of genetic structure at such a small scale requires the use of highly variable genetic markers. To this end, we describe the characterization of four highly variable microsatellite DNA markers from *P. smintheus*.

Materials and methods

Isolation of microsatellites

We isolated microsatellites with two approaches, first using bacterial transformation and clone screening (e.g., Paetkau & Strobeck 1994), and second using a method originally described by Grist *et al.* (1993), based on the polymerase chain reaction (PCR).

Bacterial transformation method

A *P. smintheus* genomic library was constructed by digesting butterfly DNA with *Sau*3A1 and ligating fragments 200-600 bp long into M13 vector mp18 which was digested with *Bam* H1 and dephosphorylated. Products of ligation were transformed into competent *E. coli* DH5α F' cells by electroporation (*E. coli Pulser*™, Bio-Rad) and transformed cells were plated onto YT medium with IPTG and X-GAL. Recombinant clones were screened using a biotinylated $(dT\text{-}dG)_{12}$ probe. Screening was done as described by Paetkau and Strobeck (1994). For clones that gave strong, positive signals, phage suspensions were amplified with modified universal forward and reverse sequencing primers (forward primer: 5'CGACGTTGTAAAACGACGGCCAG3'; reverse primer:

5'CAGGAAACAGCTATGACC3'). The PCR products were purified, and sequenced on an Applied Biosystems Inc. (ABI) 373A automated sequencer using a *Taq* DyeDeoxy™ Cycle Sequencing Kit. Primers for PCR amplification of putative microsatellites were constructed using OLIGO software (National Biosciences Inc., Version 4.0).

PCR-based method

Butterfly genomic DNA and M13 vector mp18 were ligated as above. The products of ligation were serially diluted by factors of 10 to determine the dilution at which discrete products would be produced by PCR amplification with a microsatellite-specific primer $(dT\text{-}dG)_{12}$ and a modified universal forward sequencing primer (sequence provided above). Unless otherwise stated, amplifications were in 25 µL reactions with 1X PCR buffer (10 mM Tris, pH 8.8, 0.1% Triton X100, 50 mM KCl and 0.16 mg/ml BSA), 2 mM $MgCl_2$, 0.16 µM of each primer, 120 µM dNTP, 0.3 units of *Taq* DNA polymerase (purified as described by Engelke (1990)), and 1 µL of ligation dilution as template. The cycling conditions were 1 min. at 94 °C, followed by 35 cycles of 30 sec. at 94 °C, 20 sec. at 56 °C, and 7 sec. at 72 °C, followed by 30 sec. at 72 °C, on a Perkin-Elmer Cetus 9600 thermal cycler. "Working stocks" (Grist *et al.* 1993) were made by using 1, 2, 5, and 10 µL of the appropriate ligation dilution as template in a PCR with 20 cycles and with the universal forward and reverse sequencing primers. Working stocks were diluted 100-fold and 1 µL of each dilution was used to seed a PCR amplification with the original microsatellite-specific primer and the forward sequencing primer. The products of these reactions were electrophoresed on agarose gels and individual microsatellite loci were identified as sharp bands. DNA from each band was purified, and sequenced from the forward sequencing primer using a *Taq* DyeDeoxy™ Cycle Sequencing Kit and an Applied Biosystems Inc. (ABI) 377 automated sequencer. Primers were designed from this sequence and used with the reverse sequencing primer to amplify across the microsatellite, using the corresponding working stock dilution as template. These products

were sequenced from the reverse sequencing primer, and the second, complementary primer was designed using OLIGO software (National Biosciences Inc., Version 5.0).

Sample collection and DNA isolation

Adult butterfly samples were collected from seventeen sites along Lusk Ridge (51°00'N, 114°58'W), Jumpingpound Ridge (50°57'N, 114°53'W), and Cox Hill (50°58'N, 114°54'W) in Kananaskis region, Alberta, Canada in 1995 and 1996 (Fig. 2-1). Samples were collected in conjunction with a mark-recapture study of *P. smintheus* (Roland *et al.* 2000) in which butterflies were captured with hand-nets. Throughout the course of the mark-recapture study, small (approx. 0.15 cm^2) wing clippings were taken from butterflies the first time that they were captured and marked. In 1995, such clippings were taken from all butterflies marked. In 1996 they were taken from 50% of the butterflies marked. Removing these small pieces of tissue from the wings had no effect on survival or between-meadow movement (Roland *et al.* 2000). In 1995, whole butterflies were also collected from each site on the last day that the site was visited for the mark-recapture study. All samples were placed individually in glassine envelopes and stored at -80 °C. Genomic DNA was isolated from samples using QIAmp™ spin columns (QIAGEN). For wing clippings, the entire sample was used. For whole butterflies, a sample of abdominal tissue (for males) or thoracic tissue (for females) was used.

PCR amplification and analysis of microsatellites

Loci were amplified separately in reactions containing 1X PCR buffer (10 mM Tris, pH 8.8, 0.1% Triton X100, 50 mM KCl and 0.16 mg/ml BSA), 2 mM $MgCl_2$, 0.16 μM of each primer, 120 μM dNTP, 0.3 units of *Taq* DNA polymerase, and approximately 75 ng of butterfly genomic DNA. For Ps162, the $MgCl_2$ concentration was lowered to 1.8 mM. Cycling conditions were 1 min. at 94 °C, followed by three cycles of 30 sec. at 94 °C, 20 sec. at 54 °C, and 5 sec. at 72 °C, followed by 33 cycles of 15 sec. at 94 °C, 20 sec. at 54

°C, and 1 sec. at 72 °C, followed by 30 sec. of final extension at 72 °C, on a Perkin Elmer 9600 thermal cycler. One primer of each primer set was end-labeled with a fluorescent dye, either 6-FAM, HEX, or TET (ABI). Amplification products were electrophoresed and detected using an ABI 373A Automated Sequencer and analyzed using GENESCAN and GENOTYPER software (ABI).

Landscape variables

Using topographic maps (1:50,000 scale) and 1993 black and white aerial photographs (1:40,000 scale), the following were measured as pair-wise distances between sites: total distance (meters), distance through forest (meters), distance through meadow (meters), and total elevation change (both up and down - meters) which is a measure of topography (Roland *et al.* 2000). The distance through forest and the distance through meadow necessarily sum to total distance. All distances were measured between the centroids of butterfly observations within each site (Roland *et al.* 2000). Distances between sites were not measured "as the crow flies", but as the distances that would have to be moved along the major ridges to get from one site to another. This is based on our observations that movements of *P. smintheus* are largely constrained along ridge-tops.

Estimating butterfly movements

Of the butterflies marked at any given site in 1995, the numbers that would have been recaptured at all other sites that same year were predicted from the generalized linear model with Poisson errors (Roland *et al.* 2000) that describes butterfly movements among these same meadows. These numbers were expressed as the proportion of butterflies marked in one site expected to be recaptured in the second site. These predictions were not necessarily symmetrical; the proportion of marked butterflies from site "A" recaptured in "B" would not equal the proportion of marks from "B" recaptured in "A". Therefore, for each pair of sites, the predicted proportion of marks moving from each to the other was

averaged. The predicted number of recaptures were used instead of observed numbers so that we could obtain non-zero estimates of movement for all pairs of sites, including those for which no dispersal was actually observed.

Statistical analysis

For each locus at each site, genotype frequencies were tested for conformity to Hardy-Weinberg expectations (Guo & Thompson 1992). Because we found evidence for the presence of null or non-amplifying alleles (Callen *et al.* 1993; Paetkau & Strobeck 1995; Pemberton *et al.* 1995) at all loci, for each locus at each site a maximum likelihood estimate of the frequency of the null allele was calculated, and the frequencies of all other alleles were simultaneously re-estimated, using the estimation-maximization (EM) algorithm (Ceppellini *et al.* 1955; Yasuda & Kimura 1968; Long *et al.* 1995; calculator available at <http://www.biology.ualberta.ca/jbrzusto/nullele>). This algorithm is commonly used to estimate allele frequencies in the human ABO blood group system. A null allele at a microsatellite locus is analogous to the O allele of the ABO group; it is recessive to all other alleles, and individuals that appear homozygous may either be truly homozygous or heterozygous with a null allele.

From the maximum likelihood estimates of allele frequencies, unbiased estimates of expected heterozygosity (H_e) were calculated (Nei & Roychoudhury 1974). Unbiased estimates of the frequency of heterozygotes with one null allele (H_n) were calculated as $2p_n(1-p_n)(2N/2N-1)$ where p_n is the estimated frequency of the null allele at a particular locus in a particular site, and N is the number of individuals sampled at that site. Also from the maximum likelihood estimates of allele frequencies, Nei's standard genetic distance (Nei 1972) was calculated between pairs of sites using the genetic distance calculator available at <http://www.biology.ualberta.ca/jbrzusto>. Other measures of genetic distance have been designed specifically for microsatellite loci (Goldstein *et al.* 1995, Shriver *et al.* 1995). These assume some form of step-wise mutation and incorporate information on

differences in size among alleles. We did not use such measures because we found significant size variation in the flanking region of one locus (Ps162), indicating that distances based on allele size were inappropriate for our microsatellite data set. As a measure of the amount of between-site variation relative to within-site variation, G_{st} (Nei 1973, equation 9) was calculated for all sites, using heterozygosities averaged across loci.

Genetic distances, all landscape variables, and the predicted numbers of butterflies moving between sites were expressed as pair-wise distance or similarity measures. The relationships of these measures to each other were therefore investigated using Mantel tests (Mantel 1967) and partial Mantel tests (Smouse *et al.* 1986), performed using the R-PACKAGE (Legendre & Vaudor 1991). For each test, 3000 matrix randomizations were used. The total experimentwise error for all tests was limited to 0.05 using a sequential Bonferroni adjustment (Sokal & Rohlf 1995).

Results

Isolation of microsatellites

In the bacterial transformation method, approximately 1700 recombinant clones were screened. One hundred and twenty-four clones gave a positive signal, 23 of the strongest of which were sequenced. Five of the sequenced clones contained at least five uninterrupted tandem (GT) repeats. One other clone contained a tetranucleotide repeat $(GATA)_{10}$. Only the primer set amplifying the tetranucleotide repeat (Ps50) produced clear and polymorphic amplification products, and redesigning primers for the other repeats did not improve their amplification.

The PCR-based method (Grist *et al.* 1993; Cooper *et al.* 1997) was more successful in isolating microsatellite loci from *P. smintheus*. From a total of 180 working stocks, we identified eight putative microsatellites, all from the working stocks made with 10 µL of the 10^{-4} ligation dilution as template. There was sufficient flanking sequence on both sides of

the microsatellite for primers to be designed in four cases. Originally, only one of the primer pairs (Ps162) consistently yielded clear and polymorphic amplification products. The other three primer pairs yielded products that were clearly microsatellites, but only with some samples; with many samples, no product was observed. For each of these three putative loci, we designed 3 different unlabeled and 2 different labeled primers and tested all six possible combinations. We found primer pairs that consistently amplified two of the loci (Ps81 and Ps85). Table 2-1 presents the primer sequences for each locus. Both male and female heterozygotes were typed at each locus, indicating that none of the loci are sex-linked.

When individuals were first typed at Ps162, we observed that some alleles were always in the 263-275 bp size-range, while others were in the 313-323 bp size-range. We observed no alleles in the 275-313 bp range. To ascertain that these alleles represented the same locus and to determine the origin of the size difference, we sequenced alleles of both size classes using chimeric primers consisting of an M13 primer attached to the microsatellite primer (Paetkau & Strobeck 1995). We confirmed that they were alleles of the same locus, with a 50 bp insertion/deletion near the AC-strand primer differentiating them. We also found a one bp indel and five point mutations that differed between the sequences.

Hardy-Weinberg equilibrium and evidence for null alleles

Of the 68 tests of conformity to Hardy-Weinberg proportions that were performed, 58 showed a significant departure from expected proportions. Even when the total experimentwise error rate was controlled by a Bonferroni adjustment (α=.05/68=0.0007), 32 tests were significant. Deviations were all heterozygote deficiencies, and occurred at all four loci and in every site.

At three of the four loci, Ps81, Ps85, and Ps162, we found evidence for null alleles in the form of null homozygotes. These are individuals for whom no detectable PCR

product could be observed at that particular locus, even with repeated attempts, though template from the same individual yielded products at all other loci. We observed one null homozygote at locus 81 (site I), one at locus 85 (site F), and four at locus 162 (sites D, G, I, and Y).

Although we did not observe any null homozygotes at Ps50, we found evidence for null alleles at this locus from analysis of the electropherograms produced by the gel analysis software GENOTYPER (ABI) (Fig. 2-2). The use of the electropherograms to detect null alleles is based on two expectations: (1) a band representing a particular allele in homozygous form should be approximately twice as intense as a band for that allele in heterozygous form, and (2) the ratios of the intensities of particular bands from different loci should be fairly consistent among individuals (as long as they have the same genotypes at those loci). An individual appearing homozygous at a particular allele, but with approximately half the expected band intensity, is probably a heterozygote with a null allele. For individuals having allele 106 at Ps50 and being heterozygous for alleles 124 and 126 at Ps81, the ratio of the intensity of allele 106 to the average intensity of 124 and 126 ranged from 0.81 to 1.43 with a mean of 1.14 in heterozygotes at Ps50 (n=7), and from 1.59 to 1.92 with a mean of 1.92 in most homozygotes at Ps50 (n=8). However, three individuals appearing homozygous at Ps50 had band intensity ratios of 0.84, 1.00, and 1.25, falling within the observed range for heterozygotes. These individuals probably had a null allele at Ps50.

Maximum likelihood estimation of allele frequencies

The maximum likelihood estimates of allele frequencies, including null alleles, are presented in Appendix 2-A. Averaged over all sites, the frequencies of the null alleles were 0.099 (±0.066 SD) at locus Ps50, 0.239 (±0.066 SD) at Ps81, 0.155 (±0.056 SD) at Ps85, and 0.183 (±0.064 SD) at Ps162. The mean unbiased estimate of the frequency of heterozygotes that have one null allele was 0.174 (±0.107 SD) at Ps50, 0.362 (±0.074 SD)

at Ps81, 0.260 (±0.087 SD) at Ps85, and 0.298 (±0.084 SD) at Ps162 (Table 2-2, Appendix 2-A). Averaging over all sites and loci, the unbiased estimate of expected heterozygosity was 0.707 (±0.090 SD) (Table 2-2, Appendix 2-A). Across all sites, G_{st} was 0.031.

Mantel tests and the effects of landscape structure on genetic structure

To examine the relationships among genetic distances, individual movements, and landscape variables, we performed a total of eight simple and partial Mantel tests (Table 2-3). Nei's standard genetic distance and the ln of the predicted proportion of marked animals moving between sites were significantly negatively correlated (Fig. 2-3). Genetic distance was significantly positively correlated with total geographic distance (Fig. 2-4), the distance through meadow, and the distance through forest. The correlation between genetic distance and total elevation change was not significant (Fig. 2-5).

The coefficient of determination (r^2) for genetic distance and the distance through forest was almost double that for genetic distance and the distance through meadow (Table 2-3). However, because the distance through forest and the distance through meadow were themselves highly correlated, we used partial Mantel tests to tease apart the effects of these two landscape features. The partial correlation of genetic distance and the distance through forest, controlling for the distance through meadow, was not significant at the Bonferroni α level; however, the correlation coefficient was fairly high ($r = 0.343$) and the probability value was low ($P = 0.04$). The partial correlation of genetic distance and the distance through meadow, controlling for the distance through forest, was also not significant; the correlation coefficient was extremely low ($r = -0.001$) and the probability value was high ($P = 0.49$). The coefficient of multiple determination (R^2) for these two variables was 0.222 which was equal to the coefficient of determination between genetic distance and the distance through forest. Thus, having fitted the distance through forest, no further variation in genetic distance was accounted for by adding the distance through

meadow. This pattern is evident in the relationship between Nei's standard genetic distance and the residuals derived from a regression of the distance through meadow on the distance through forest (Fig. 2-6a); once the distance through forest is accounted for, there is no relationship between genetic distance and the distance through meadow. However, having fit the distance through meadow, by adding the distance through forest, the variation in genetic distance that was explained doubled. This is seen in the relationship between Nei's standard genetic distance and the residuals derived from a regression of the distance through forest on the distance through meadow (Fig. 2-6b); there is still a positive relationship between genetic distance and the distance through forest even when the distance through meadow has been accounted for.

The above analysis was repeated using pairwise estimates of Nei's standard genetic distance that were calculated using the observed allele frequencies and not accounting for the presence of null alleles. The patterns of association among variables and conclusions reached were the same as those presented in Table 2-3 (data not shown).

Analysis restricted to Jumpingpound Ridge and Cox Hill

Because we were concerned about the extent to which our results were determined by the large valley separating Lusk Ridge from Jumpingpound Ridge (Fig. 2-1), we repeated the Mantel tests removing sites D and E. Genetic distance was still significantly correlated with predicted dispersal (r = -0.409, P = 0.001), total distance (r = 0.384, P = 0.00067), forest distance (r = 0.355, P = 0.012), and meadow distance (r = 0.361, P = 0.00067). At this scale, forest distance and meadow distance explained equal amounts of variation in genetic distance. The correlation coefficients for genetic distance with each of forest distance and meadow distance were approximately equal, and the partial correlation of genetic distance with forest distance, controlling for meadow distance (r = 0.152, P = 0.213) was similar to the partial correlation of genetic distance with meadow distance, controlling for forest distance (r = 0.166, P = 0.072). However, on Jumpingpound Ridge, there is much more

meadow than forest (Fig. 2-1) and a similar overall effect of both habitat types suggests that per unit of forest there is a greater reduction of genetic distance than per unit of meadow. This is confirmed by comparison of the slopes of linear regressions of genetic distance against forest distance ($b = 0.017$) and genetic distance against meadow distance ($b = 0.007$). Also, when the analysis was repeated using only the ten sites F-K, Y, and Z, among which there is a more even distribution of the two habitat types, the relative values of correlation coefficients were the same as those presented in Table 2-3, although none of the tests were significant because of a lack of power (data not shown).

Discussion

Microsatellites in butterflies

The development of microsatellite DNA markers in butterflies is more difficult than in many other species (Meglécz & Solignac 1998). Difficulty developing microsatellites could result from their low frequency or from some factor that impedes successful amplification. We do not know if or why microsatellites might be rare in butterflies. However, several aspects of our results suggest a high rate of mutation in the regions immediately surrounding microsatellite repeats in butterflies, which would make primer design and PCR amplification challenging. First, we observed a high frequency of null alleles at all four loci. Most null alleles arise from mutations in the primer-binding sites (Callen *et al* 1993: Paetkau & Strobeck 1995). Second, for two loci (Ps81 and Ps85) our original primer pairs amplified what appeared to be microsatellites, but we had to redesign primers before amplification products could be obtained from most samples. Apparently, the original primer pairs suffered from a higher frequency of null alleles than the latter ones. Third, when alleles of locus Ps162 were sequenced, both point mutations and insertions/deletions were detected in the regions surrounding the microsatellite repeat.

Two other studies have described the characterization of microsatellites in butterfly species: Palo *et al.* (1995) reported two microsatellites for *Melitaea cinxia*, and Meglécz *et al.* (1998) and Meglécz & Solignac (1998) reported three for *Parnassius mnemosyne*. Both reported excess homozygosity at all loci. Palo *et al.* (1995) argued that population structure cannot account for the large deviations from Hardy-Weinberg proportions. Their results did not ultimately reveal the mechanism producing the observed deviations, but null alleles are a possible explanation. Although Meglécz & Solignac (1998) argued that it was unlikely that null alleles were responsible for the excess homozygosity they observed, they could not exclude the possibility that null alleles existed. In addition, Palo *et al.* (1995) reported that sizes of alleles of their dinucleotide loci were not always integral multiples of two and that in sequencing some alleles, along with variation in repeat number, they found other types of sequence rearrangements. All these results are consistent with the hypothesis that there is high instability in the microsatellite-flanking regions in butterfly species.

Therefore, to develop butterfly microsatellites, we suggest that effort be invested in designing and testing a number of primer pair combinations for each putative locus. It is also important to be wary of the high probability that there will be null alleles. If null alleles are common, their frequency can be estimated and they can be treated as an extra allelic class, as we have done. A more rigorous, but time-consuming, option would be to use several different primer sets for every locus, that together will amplify most alleles.

Dispersal rates and genetic distances

The significant correlation between genetic distance and predicted dispersal indicates that, in our study area, gene flow results largely from the types of movement events observed using mark-recapture (Table 2-3, Fig. 2-3). In a number of population-genetic studies, patterns of genetic variation suggested more gene flow than would be expected given known dispersal distances (e.g., Cullenward *et al.* 1979, Peterson 1996). In such cases, it

is hypothesized that gene flow occurs via occasional episodes of long-distance dispersal, or from extinction-colonization dynamics (Slatkin, 1985), and not from the types of movement events that are readily observable. These other mechanisms of genetic exchange need not be invoked here. However, it should be noted that gene flow among more distant sites is likely to be via stepping-stones rather than direct movements. Also, because gene flow depends on successful breeding as well as dispersal (Gilbert & Singer 1973; Slatkin 1985), this correlation suggests that individuals are contributing to the gene pool of populations into which they immigrate.

Variation in genetic distances was greater at low levels of dispersal than at high levels (Fig. 2-3). This probably reflects the fact that at low levels of dispersal, genetic drift, which is an inherently stochastic process, is the main evolutionary force affecting allele frequencies. At higher levels of dispersal, gene flow, which is a deterministic process, is the dominant evolutionary force.

Genetic structure at a small spatial scale

We were able to detect genetic structure at a small spatial scale, with only 12 km separating the most distant populations. The significant correlation between genetic distance and geographic distance (Fig. 2-4) suggests a pattern of isolation by distance, which was detectable at the even smaller scale of Jumpingpound Ridge and Cox Hill only (approximately 7 km separating the most distant populations). Clearly, rates of gene flow in the study area were not so high as to render it one genetically homogeneous unit. Our results are in contrast to those of Cullenward *et al.* (1979), Baughman *et al.* (1990), and Peterson (1995) who found very little genetic structure among butterfly populations at larger scales of thirty to hundreds of km. Two factors probably facilitated the detection of genetic structure in this study. First, we used highly variable microsatellite markers which have a higher "resolving power" than allozymes used in the other studies and are more likely to be free of selection pressures. Second, meadows in our study area are distributed

in a one-dimensional manner, which is more conducive to the detection of isolation by distance than would be a two-dimensional habitat (Slatkin, 1993). However, the discrepancy between our results and those mentioned above may also reflect the interplay between patchy habitat distribution and low vagility of *P. smintheus* through forested areas (Roland *et al.* 2000) in reducing rates of gene flow as compared to some other butterfly species.

Landscape structure and population genetic structure

We found that the distance through forest separating populations was a much better predictor of genetic dissimilarity than was distance through meadow (Fig. 2-6, Table 2-3). Therefore, landscape structure had a strong influence on the genetic differentiation of populations. Roland *et al.* (2000) found that both distance through meadow and distance through forest affected dispersal between sites, but the rate at which movements were reduced per km of forest was much greater than the rate at which movements were reduced per km of meadow. The agreement of our results with those of Roland *et al.* (2000) and the significant correlation between genetic distance and dispersal rate suggest that the correlations between landscape variables and genetic distance largely reflect the effects of landscape structure on dispersal. Meglécz *et al.* (1997) suggested that forests are also barriers to gene flow for *P. mnemosyne*.

We did not find a significant correlation between elevation change and genetic distance (Table 2-3, Fig. 2-5). Roland *et al* (2000) found a significant effect of elevation change on dispersal, although this effect was weak compared to those of distance through forest and distance through meadow. Topography may influence dispersal rates, but this effect was not reflected significantly in the geographic distribution of genetic variation.

We found effects of landscape on population genetic structure concurrent with a significant pattern of isolation by distance (Fig. 2-4, Table 2-3). Britten *et al.* (1995) and Johnson & Black (1995) found isolation by distance in regions where habitat was

continuous and no isolation by distance in regions that were highly fragmented. Their results suggest that when landscape features other than geographical distance influence population genetic structure, isolation by distance is not detected. Our results show, however, that the occurrence of isolation by distance does not necessarily rule out the possibility that features of the physical landscape other than geographical distance influence patterns of genetic differentiation.

Implications for long-term persistence

P. smintheus is abundant and widespread in North America. However, the congeners *P. mnemosyne* and *P. apollo* are threatened in Europe (Väisänen & Somerma 1985; Bengtsson *et al.* 1989, Witkowski *et al.* 1997). In Europe, the decline of *P. apollo* is partly attributed to changes in habitat structure; in particular, to a reduction in the area of open, meadow habitat and an increase in forest cover (Witkowski *et al.* 1997). In North American alpine areas, climatic warming may be causing a similar rise in tree-line and a reduction in the amount of high-altitude, meadow habitat (Kearney 1982; Taylor 1995; Woodward *et al.* 1995). Our results indicate that such a change in the landscape will lead to reduced genetic exchange among populations of *P. smintheus*. Increasing isolation will probably lead to reduced genetic variation within populations, irrespective of changes in population size, because genetic drift will not be as strongly countered by gene flow. In a study of *P. mnemosyne* populations, Descimon & Napolitano (1993) found a significant negative correlation between population isolation and genetic variability. Reduced genetic variation will, in turn, have implications for population survival (Saccheri *et al.* 1998) and for the ability of populations to adapt to changing environmental conditions in the long term (Templeton *et al.* 1990). Therefore, *P. smintheus* is vulnerable to changes in landscape structure and the large-scale encroachment of forests into alpine meadow habitat represents a potential threat this species, in at least part of its range.

Literature cited

Arter HE (1990) Spatial relationship and gene flow paths between populations of the alpine snail *Arianta arbustorum* (Pulmonata: Helicidae). *Evolution* , **44**, 966-980.

Avise JC, Felley J (1979) Population structure of freshwater fishes I. Genetic variation of bluegill (*Lepomis macrochirus*) populations in man-made reservoirs. *Evolution*, **33**, 15-26.

Baughman JF, Brussard PF, Ehrlich PR, Murphy DD (1990) History, selection, drift, and gene flow: complex differentiation in checkerspot butterflies. *Canadian Journal of Zoology*, **68**, 1967-1975.

Bengtsson BE, Elmquist H, Nyholm E (1989) On the Swedish apollo butterfly with an attempt to explain its decline. *Entomologisk Tidskrift*, **110**, 31-37.

Britten HB, Brussard PF, Murphy DD, Ehrlich PR (1995) A test for isolation-by-distance in Central Rocky Mountain and Great Basin populations of Edith's Checkerspot butterfly (*Euphydryas editha*). *Journal of Heredity* , **86**, 204-210.

Caccone A, Sbordoni V (1987) Molecular evolutionary divergence among North American cave crickets. I. Allozyme variation. *Evolution*, **41**, 1198-1214.

Callen DF, Thompson AD, Shen Y, *et al.* (1993) Incidence and origin of null alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics*, **52**, 922-927.

Ceppellini R, Siniscalco M, Smith CAB (1955) The estimation of gene frequencies in a random-mating population. *Annals of Human Genetics* , **20**, 97-115.

Cooper SJB, Bull CM, Gardner MG (1997) Characterization of microsatellite loci from the socially monogamous lizard *Tiliqua rugosa* using a PCR-based isolation technique. *Molecular Ecology* , **6**, 793-795.

Cullenward MJ, Ehrlich PR, White RR, Holdren CE (1979) The ecology and population genetics of an alpine checkerspot butterfly, *Euphydryas anicia*. *Oecologia* , **38**, 1-12.

Descimon H, Napolitano M (1993) Enzyme polymorphism, wing pattern variability, and geographical isolation in an endangered butterfly species. *Biological Conservation* , **66**, 117-123.

Engelke DR, Krikos A, Bruck ME, Ginsburg D (1990) Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. *Analytical Biochemistry*, **191**, 396-400.

Fisher RA, Ford EB (1947) The spread of a gene in natural conditions in a colony of moth *Panaxia dominula* L. *Heredity*, **1**, 143-174.

Gilbert LE, Singer MC (1973) Dispersal and gene flow in a butterfly species. *American Naturalist*, **107**, 58-72.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences USA* **92**: 6723-6727.

Grist SA, Firgaira FA, Morley AA (1993) Dinucleotide repeat polymorphisms isolated by the polymerase chain reaction. *BioTechniques*, **15**, 304-309.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361-372.

Johannesen J, Veith M, Seitz A (1996) Population genetic structure of the butterfly *Melitaea didyma* (Nymphalidae) along a northern distribution range border. *Molecular Ecology*, **5**, 259-267.

Johnson MS, Black R (1995) Neighbourhood size and the importance of barriers to gene flow in an intertidal snail. *Heredity*, **75**, 142-154.

Kane TC, Barr TC, Badaracca WJ (1992) Cave beetle genetics: geology and gene flow. *Heredity*, **68**, 277-286.

Kearney MS (1982) Recent seedling establishment at timberline in Jasper National Park, Alta. *Canadian Journal of Botany*, **60**, 2283-2287.

King PS (1987) Macro- and microgeographic structure of a spatially subdivided beetle species in nature. *Evolution*, **41**, 401-416.

Legendre P, Vaudor A (1991) *The R package: multidimensional analysis, spatial analysis.* Version 3.0. Universite de Montréal, Montréal.

Long JC, Williams RC, Urbanek M (1995) An E-M algorithm and testing strategy for multiple-locus haplotypes. *American Journal of Human Genetics*, **56**, 799-810.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209-220.

Mayr E (1942) *Systematics and the Origin of Species.* Columbia University Press, New York.

Meglécz E, Pecsenye K, Peregovits L, Varga Z (1997) Effects of population size and habitat fragmentation on the genetic variability of *Parnassius mnemosyne* populations in NE Hungary. *Acta Zoologica Academiae Scientiarum Hungaricae*, **43**, 183-190.

Meglécz E, Pecsenye K, Varga Z, Solignac M (1998) Comparison of differentiation pattern at allozyme and microsatellite loci in *Parnassius mnemosyne* (Lepidoptera) populations. *Hereditas*, **128**, 95-103.

Meglécz E, Solignac M (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas*, **128**, 179-180.

Napolitano M, Descimon H (1994) Genetic structure of French populations of the mountain butterfly *Parnassius mnemosyne* L. (Lepidoptera: Papilionidae). *Biological Journal of the Linnean Society*, **53**, 325-341.

Nei M (1972) Genetic distance between populations. *American Naturalist*, **106**, 283-292.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, **70**, 3321-3323.

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. *Genetics*, **76**, 379-390.

Nève G, Mousson L, Baguette M (1996) Adult dispersal and genetic structure of butterfly populations in a fragmented landscape. *Acta Oecologia*, **17**, 621-626.

Paetkau D, Strobeck C (1994) Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology*, **3**, 489-495.

Paetkau D, Strobeck C (1995) The molecular basis and phylogenetic distribution of a microsatellite null allele in bears. *Molecular Ecology*, **4**, 519-520

Palo J, Varvio S, Hanski I, and Vainola R. (1995) Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. *Hereditas*, **123**, 295-300.

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*, **4**, 249-252.

Peterson MA (1995) Phenological isolation, gene flow, and developmental differences among low- and high-elevation populations of *Euphilotes enoptes* (Lepidoptera: Lycaenidae). *Evolution*, **49**, 446-455.

Peterson MA (1996) Long-distance gene flow in the sedentary butterfly, *Euphilotes enoptes* (Lepidoptera: Lycaenidae). *Evolution*, **50**, 1990-1999.

Pounds JA, Jackson JF (1981) Riverine barriers to gene flow and the differentiation of fence lizard populations. *Evolution*, **35**, 516-528.

Preziosi RF, Fairbairn DJ (1992) Genetic population structure and levels of gene flow in the stream dwelling waterstrider, *Aquarius* (=*Gerris*) *remigis* (Hemiptera: Gerridae). *Evolution*, **46**, 430-444.

Roland, J., N. Keyghobadi, and S. Fownes (2000) Alpine *Parnassius* butterfly dispersal: effects of landscape and population size. *Ecology* **81**: 1642-1653.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature*, **392**, 491-494.

Scott JA (1973) Population biology and adult behavior of the circumpolar butterfly *Parnassius phoebus* F. (Papilionidae). *Entomologia scandinavia*, **4**, 161-168.

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution* **12**: 914-920.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics*, **16**, 393-430.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* , **47**, 264-279.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627-632.

Sokal RR, Rohlf FJ (1995) *Biometry*, 3rd edn. W. H. Freeman and Company, New York, NY.

Sperling FAH, Kondla NG (1991) Alberta Swallowtails and Parnassians: Natural history, keys, and distribution. *Blue Jay*, **49**, 183-192.

Taylor AH (1995) Forest expansion and climate change in the mountain hemlock (*Tsuga mertensiana*) zone, Lassen Volcanic National Park, California, USA. *Arctic and Alpine Research*, **27**, 207-216.

Templeton AR, Shaw K, Routman E, Davis SK (1990) The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden*, **77**, 13-27.

Väisänen R, Somerma P (1985) The status of *Parnassius mnemosyne* (Lepidoptera: Papilionidae) in Finland. *Notulae Entomologicae*, **65**, 109-118.

VanDongen S, Backeljau T, Matthysen E, Dhondt AA (1998) Genetic population structure of the winter moth (*Operophtera brumata* L.) (Lepidoptera, Geometridae) in a fragmented landscape. *Heredity*, **80**, 92-100.

Witkowski Z, Adamski P, Kosior A, Plonka P (1997) Extinction and reintroduction of *Parnassius apollo* in the Pieniny National Park (Polish Carpathians). *Biologia (Bratislava)*, **52**, 199-208.

Woodward A, Schreiner EG, Silsbee DG (1995) Climate, geography, and tree establishment in subalpine meadows of the Olympic Mountains, Washington, U.S.A. *Arctic and Alpine Research*, **27**, 217-225.

Yasuda N, Kimura M (1968) A gene-counting method of maximum likelihood for estimating gene frequencies in ABO and ABO-like systems. *Annals of Human Genetics* , **31**, 409-420.

Table 2-1. Primer sequences for four *P. smintheus* microsatellite loci. FAM, HEX, and TET are fluorescent-dye labels (AppliedBiosystems). Size range refers to the observed sizes of PCR products. Number of alleles was determined from the entire sample of 557 individuals and does not include null alleles.

Locus	Repeat motif	Primer sequences (5'to 3')	Size range (bp)	No. of alleles	Accession #
Ps50	$(GATA)_{10}$	FAM GAT CAC CGA GAA AGA GAA AG	98-126	8	AF133661
		TTT TTG CGT CTG TTA CAT AA			
Ps81	$(GT)_{10}G(GT)_2$	TET AAA TGG AGC AAT TAT ACC TA	122-136	7	AF133662
		GTT GCC CCG TTG AGT GAA AG			
Ps85	$(GT)_{10}$	HEX CAC GCT CTG GCA CTA TCT ACC	118-135	8	AF133663
		TGC GCA GAT AGG GCT GAC			
Ps162	$(AT)_2(GT)_6$	FAM GCT ACG TGC ATT GTG TTG	264-275, 315-323	12	AF133664
		GGT CGG TTT CAT TTT TGT AG			

Table 2-2. Genetic variation within each sampling site. For each site, number of alleles and heterozygosity values were averaged across the four loci. Number of alleles includes null alleles. Expected heterozygosity was calculated as $(1-\Sigma p_i^2)*(2N/2N-1)$ where p_i is the frequency of each allele and N is the number of individuals sampled. The frequency of heterozygotes with one null allele was calculated as $2p_n(1-p_n)(2N/2N-1)$ where p_n is the estimated frequency of the null allele, and N is the number of individuals sampled.

Site	Sample size	Mean number of alleles	Mean expected heterozygosity	Mean frequency of heterozygotes with one null allele
D	43	6.50	0.6439	0.2555
E	40	7.25	0.6942	0.2118
F	41	7.25	0.7571	0.2838
G1	40	6.50	0.6914	0.2381
G2	40	7.00	0.7169	0.2630
I	21	6.00	0.7110	0.2722
J	31	6.25	0.6959	0.2686
K	40	6.50	0.7110	0.2542
L	40	7.00	0.7466	0.2782
M	38	6.75	0.6878	0.2638
O	12	6.00	0.7569	0.3225
P	39	6.75	0.7158	0.2768
Q	40	7.25	0.7005	0.3007
R	24	6.50	0.7129	0.2951
S	14	5.25	0.6392	0.2655
Y	13	5.50	0.7441	0.3295
Z	41	6.25	0.6895	0.2687

Table 2-3. Results of simple and partial Mantel tests investigating the relationships among genetic distances, individual movements, and landscape variables. The first six tests are simple Mantel tests. The last two tests are partial Mantel tests: "(genXfor).mdw" tests the partial correlation between genetic distance and distance through forest, controlling for distance through meadow and "(genXmdw).for" tests the partial correlation between genetic distance and distance through meadow, controlling for distance through forest. The α value for each test was determined by a sequential Bonferroni adjustment. "*" indicates a significant test and "NS" indicates a non-significant test. "*r*" is the standardized Mantel statistic which is equivalent to a Pearson product-moment correlation coefficient and "r^2" is the coefficient of determination.

Mantel test	Observed *P* value	Bonferroni α value	Significance	*r*	r^2
meadow distanceX forest distance	0.00033	0.006391	*	0.73179	0.53552
genetic distance X ln (predicted proportion moving)	0.00167	0.0073	*	-0.44300	0.19290
genetic distance X total distance	0.00566	0.008512	*	0.42987	0.18479
genetic distance X meadow distance	0.00666	0.0102	*	0.34385	0.11823
genetic distance X forest distance	0.01066	0.01274	*	0.47087	0.22172
genetic distance X elevation change	0.01866	0.016952	NS	0.45917	0.21084
(gen X for).mdw	0.04099	-	NS	0.34258	0.11736
(gen X mdw).for	0.49317	-	NS	-0.00121	1.464E-06

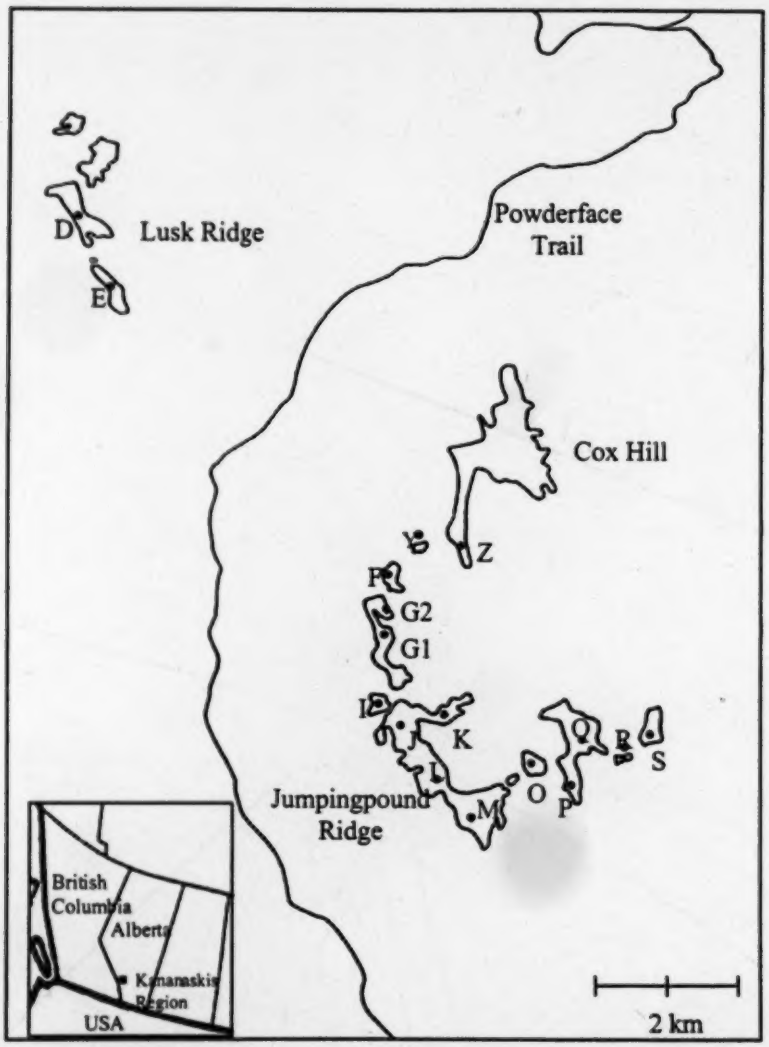
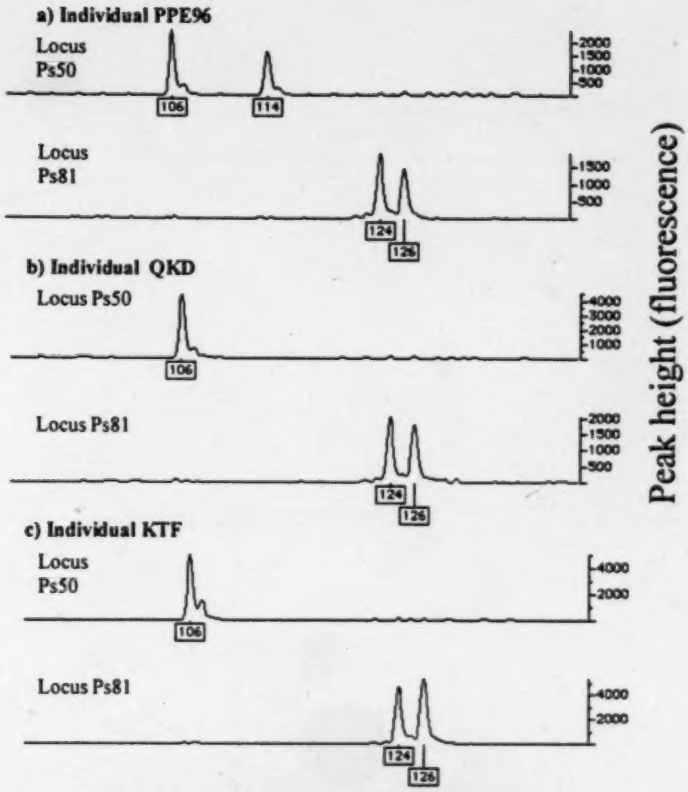


Figure 2-1. Locations of sampling sites. Each site is identified by a different letter. Outlined areas denote high altitude (> 2000m), meadow habitat.



Individual	Locus	Fragment size (bp)	Peak height	Peak area (total fluorescence)
PPE96	Ps50	106	2445	14308
	Ps50	114	1636	10014
	Ps81	124	1889	13995
	Ps81	126	1422	10606
QKD	Ps50	106	4534	27732
	Ps81	124	2032	13772
	Ps81	126	1789	14076
KTF	Ps50	106	5027	30737
	Ps81	124	4687	32221
	Ps81	126	5268	40852

Figure 2-2. Evidence for null alleles at locus Ps50 from analysis of electropherograms. For illustrative purposes, electropherograms from one individual of each of the following classes are shown: (1) heterozygote at Ps50 with ratio of intensity of band 106 at Ps50 to average of bands 124 and 126 at Ps81 between 0.81 and 1.43 (PPE96), (2) homozygote at Ps50 with band-intensity ratio between 1.59 and 2.37 (QKD), (3) homozygote at Ps50 with band-intensity ratio of 0.84 falling within the range of heterozygotes (KTF). Individual KTF likely has a null allele.

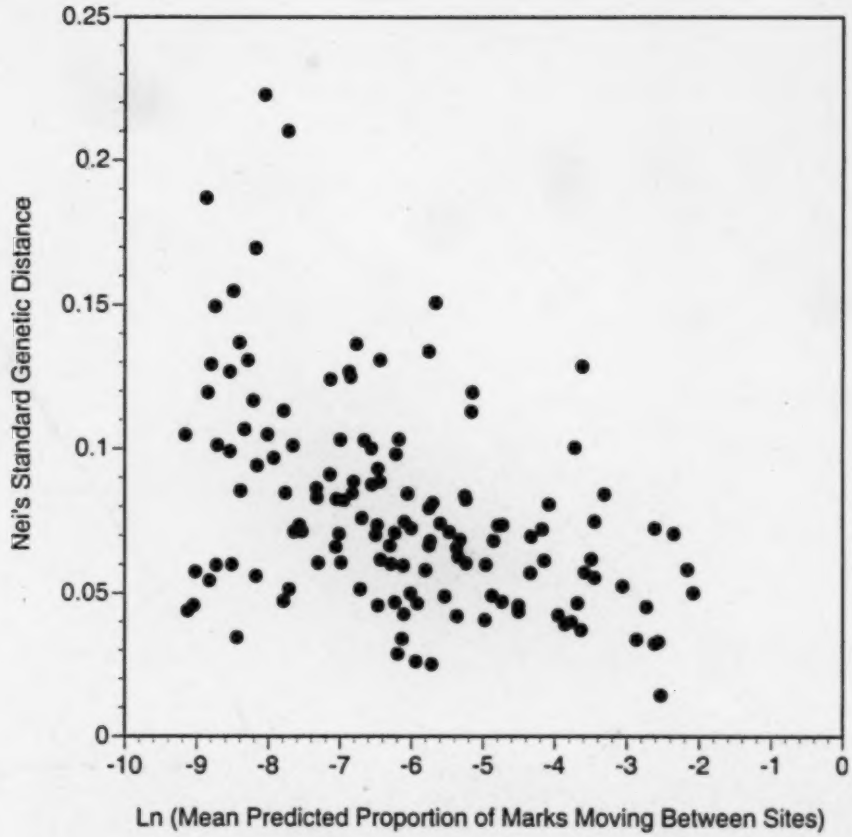


Figure 2-3. Nei's standard genetic distance versus ln of the average predicted proportion of marked butterflies moving between a pair of sites, for all pairs of sites.

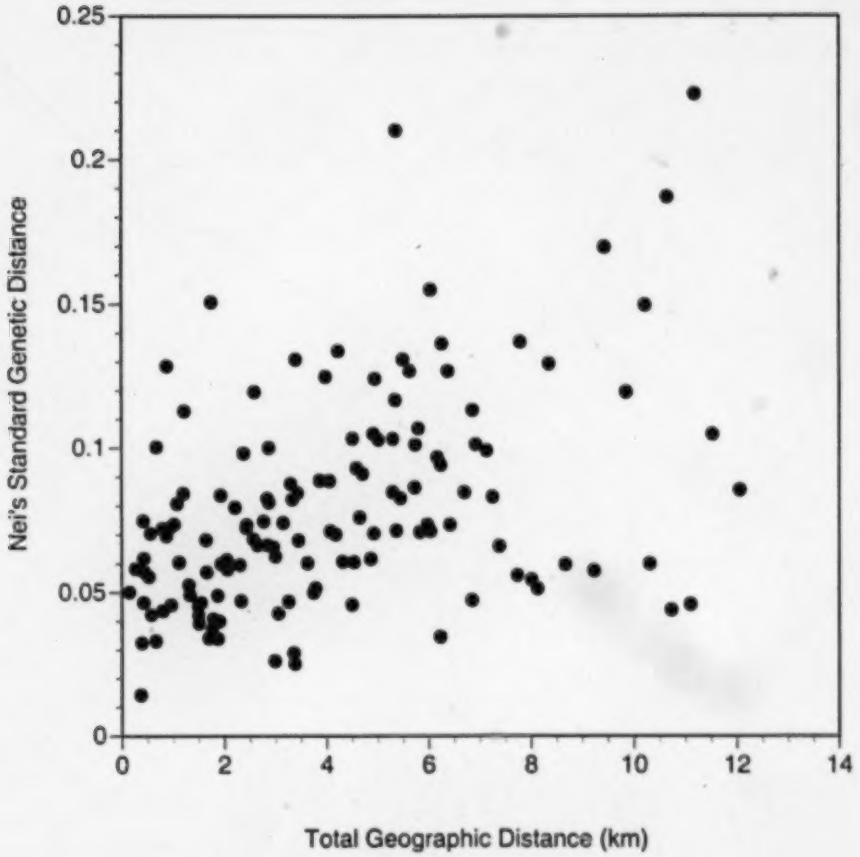


Figure 2-4. Nei's standard genetic distance versus total geographic distance, for all pairs of sites.

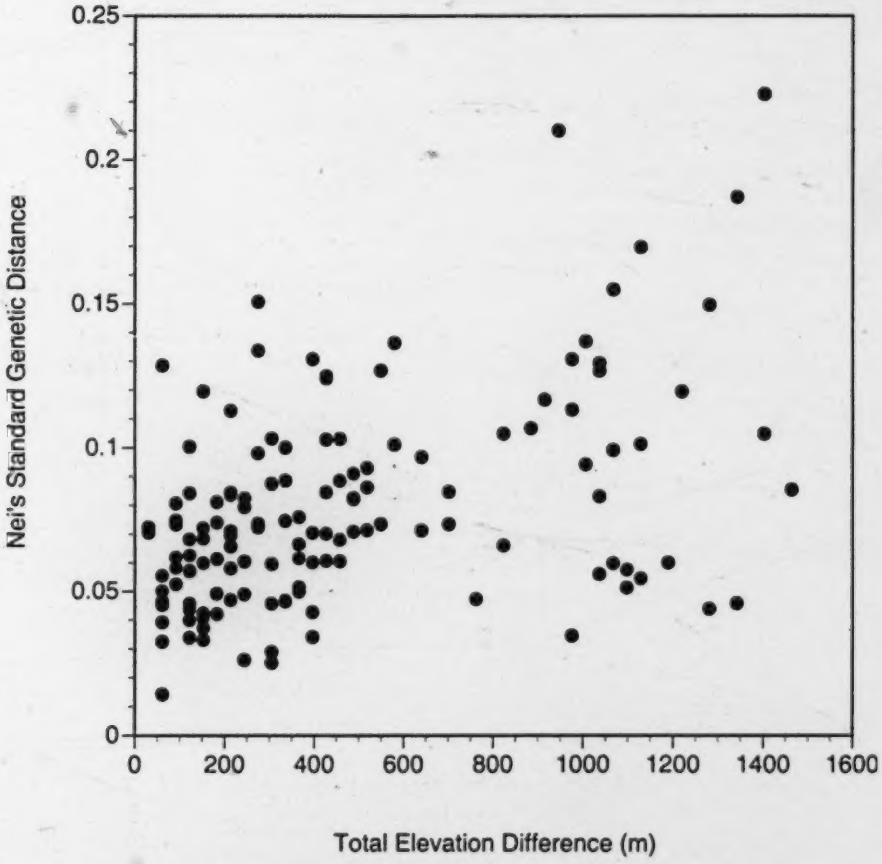


Figure 2-5. Nei's standard genetic distance versus total elevation difference, for all pairs of sites.

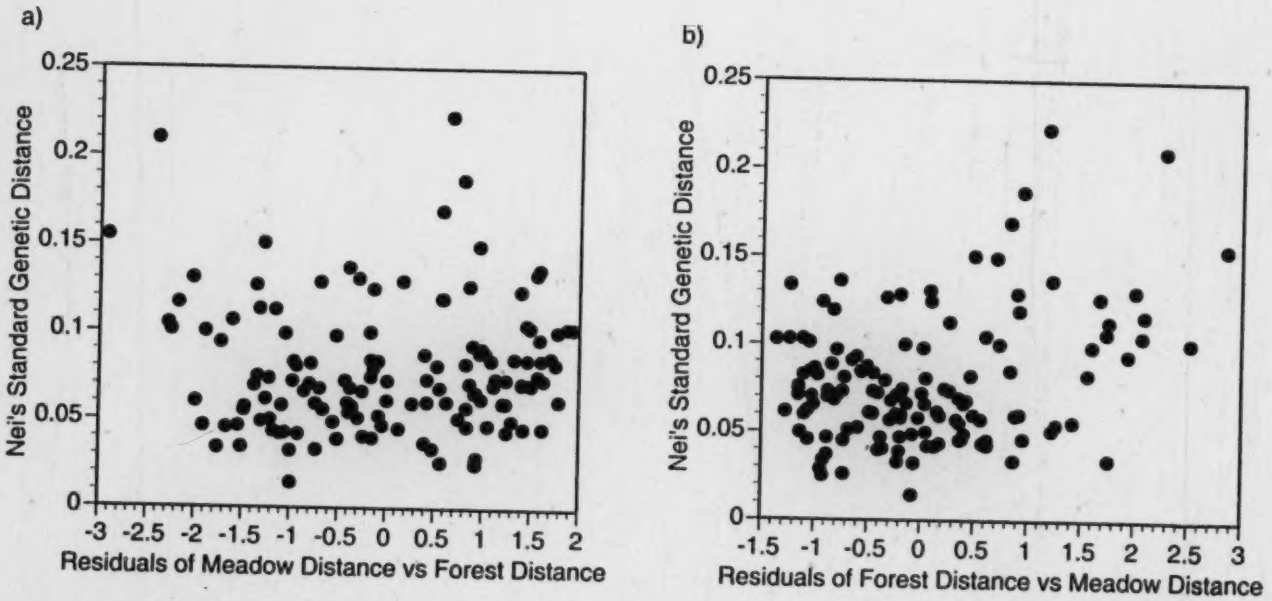


Figure 2-6. a) Nei's standard genetic distance versus residuals from a least squares linear regression of distance through meadow against distance through forest. The corresponding partial Mantel test is not significant and has a very low $r = -0.00121$. b) Nei's standard genetic distance versus residuals from a least squares linear regression of distance through forest against distance through meadow. The corresponding partial Mantel test is not significant but has a high $r = 0.34258$. Data are for all pairs of sites.

Appendix 2-A. Estimated allele frequencies by locus and population. Expected heterozygosity (H_e) and frequency of null heterozygotes (H_n) calculated as in Table 2-2.

site	locus	N	null allele	allele 1	allele 2	allele 3	allele 4	allele 5	allele 6	allele 7	allele 8	allele 9	allele 10	allele 11	allele 12	He	Hn
D	162	43	0.2	0	0.45	0.06	0.16	0	0	0.01	0	0.02	0.01	0.09	0	0.73	0.32
E	162	40	0.07	0	0.57	0.03	0.13	0.01	0	0	0.03	0.03	0.03	0.1	0.01	0.65	0.14
F	162	41	0.24	0.01	0.41	0.05	0.08	0	0	0	0	0	0.13	0.05	0.02	0.75	0.36
G1	162	40	0.21	0	0.49	0.01	0.22	0	0	0.03	0	0.01	0.01	0.01	0	0.67	0.34
G2	162	40	0.22	0	0.59	0.03	0.09	0.01	0	0.01	0	0.01	0.01	0.03	0	0.61	0.34
I	162	21	0.22	0	0.51	0.15	0.1	0	0	0.03	0	0	0	0	0	0.68	0.35
J	162	31	0.13	0	0.49	0.12	0.08	0	0	0.12	0	0	0.05	0	0	0.72	0.24
K	162	40	0.09	0	0.51	0.12	0.19	0.01	0	0.04	0	0	0.03	0	0.01	0.69	0.17
L	162	40	0.28	0	0.31	0.17	0.1	0	0	0.05	0	0.04	0.03	0.03	0	0.79	0.40
M	162	38	0.1	0	0.58	0.15	0.07	0.01	0.03	0.04	0	0	0.01	0	0	0.63	0.19
O	162	12	0.22	0	0.41	0.08	0.15	0	0	0.09	0	0	0.05	0	0	0.77	0.36
P	162	39	0.19	0	0.42	0.13	0.13	0	0	0.07	0	0	0.06	0	0	0.75	0.32
Q	162	40	0.14	0	0.41	0.18	0.16	0	0.03	0.04	0	0	0.04	0.01	0	0.76	0.24
R	162	24	0.16	0	0.39	0.04	0.18	0	0.02	0.1	0	0.02	0.08	0	0	0.78	0.27
S	162	14	0.18	0	0.52	0.07	0.12	0.04	0.04	0.04	0	0	0	0	0	0.70	0.30
Y	162	13	0.31	0	0.36	0.08	0.09	0	0	0.04	0	0	0.08	0.04	0	0.78	0.45
Z	162	41	0.15	0	0.54	0.03	0.17	0	0	0.02	0	0	0.03	0.07	0	0.66	0.26
D	50	43	0.11	0	0	0.29	0.25	0.18	0.1	0.05	0.02					0.81	0.19
E	50	40	0.01	0	0.05	0.27	0.15	0.27	0.16	0.04	0.05					0.81	0.01
F	50	41	0.01	0	0	0.34	0.26	0.23	0.12	0.02	0.01					0.76	0.03
G1	50	40	0.02	0	0	0.36	0.18	0.18	0.12	0.08	0.05					0.79	0.05
G2	50	40	0.06	0	0	0.37	0.17	0.2	0.1	0.05	0.05					0.78	0.11
I	50	21	0.04	0	0.03	0.3	0.23	0.26	0.1	0.05	0					0.80	0.07
J	50	31	0.11	0	0	0.44	0.15	0.17	0.1	0.02	0.02					0.74	0.19
K	50	40	0.08	0.01	0.03	0.19	0.22	0.28	0.12	0	0.07					0.82	0.15
L	50	40	0.08	0	0.04	0.23	0.26	0.12	0.06	0.05	0.16					0.83	0.15
M	50	38	0.13	0	0.01	0.21	0.31	0.14	0.08	0.08	0.04					0.82	0.23
O	50	12	0.19	0.09	0.04	0.29	0.26	0.08	0.05	0	0					0.83	0.33
P	50	39	0.11	0.04	0.01	0.27	0.13	0.25	0.11	0.06	0					0.82	0.21
Q	50	40	0.2	0.04	0.03	0.31	0.12	0.17	0.12	0.01	0.01					0.82	0.32
R	50	24	0.17	0	0	0.32	0.24	0.15	0.09	0.02	0					0.79	0.29
S	50	14	0.22	0	0.04	0.33	0.11	0.19	0.08	0.04	0					0.82	0.35
Y	50	13	0.1	0	0	0.19	0.27	0.32	0.08	0	0.04					0.80	0.19
Z	50	41	0.05	0	0	0.44	0.2	0.1	0.11	0.06	0.04					0.74	0.09

Appendix 2-A continued.

D	81	43	0.12	0.02	0.75	0.09	0	0.01	0.01	0		0.43	0.21
E	81	40	0.27	0.05	0.55	0.03	0.08	0	0.03	0		0.63	0.40
F	81	41	0.26	0.15	0.35	0.15	0.03	0	0.07	0		0.77	0.39
GI	81	40	0.18	0.09	0.56	0.15	0.01	0	0.01	0		0.64	0.29
G2	81	40	0.26	0.21	0.3	0.16	0.04	0	0.04	0		0.78	0.38
I	81	21	0.27	0.02	0.52	0.19	0	0	0	0		0.64	0.40
J	81	31	0.23	0.06	0.51	0.15	0.02	0	0.03	0		0.67	0.36
K	81	40	0.25	0	0.53	0.15	0.04	0	0.03	0		0.64	0.38
L	81	40	0.18	0.09	0.52	0.16	0.01	0	0.03	0		0.66	0.30
M	81	38	0.26	0.03	0.5	0.14	0	0	0.07	0		0.67	0.39
O	81	12	0.11	0.04	0.55	0.25	0.04	0	0	0		0.64	0.20
P	81	39	0.23	0.07	0.51	0.14	0.01	0	0.03	0.01		0.67	0.36
Q	81	40	0.23	0.06	0.54	0.11	0.04	0	0.01	0		0.64	0.36
R	81	24	0.36	0.04	0.36	0.17	0.04	0	0.02	0		0.72	0.47
S	81	14	0.26	0.16	0.5	0.08	0	0	0	0		0.68	0.40
Y	81	13	0.25	0.17	0.41	0.12	0	0	0.04	0		0.75	0.39
Z	81	41	0.35	0.09	0.36	0.11	0.01	0	0.08	0		0.73	0.46
D	85	43	0.18	0.12	0.01	0	0	0.58	0.11	0	0	0.61	0.30
E	85	40	0.18	0.43	0	0	0.03	0.29	0.07	0	0	0.70	0.30
F	85	41	0.22	0.19	0	0.01	0.02	0.4	0.13	0.01	0.01	0.75	0.35
GI	85	40	0.16	0.1	0	0	0.01	0.51	0.22	0	0	0.67	0.27
G2	85	40	0.12	0.14	0.01	0	0.04	0.49	0.2	0	0	0.69	0.22
I	85	21	0.15	0.18	0.02	0	0.05	0.46	0.08	0.02	0.02	0.74	0.26
J	85	31	0.17	0.16	0	0.02	0.02	0.55	0.09	0	0	0.65	0.29
K	85	40	0.19	0.11	0	0	0.03	0.47	0.21	0	0	0.70	0.31
L	85	40	0.15	0.26	0.03	0	0.01	0.46	0.09	0	0	0.70	0.26
M	85	38	0.14	0.13	0.01	0	0.04	0.57	0.1	0	0	0.63	0.25
O	85	12	0.26	0.13	0.1	0	0.04	0.38	0.09	0	0	0.78	0.40
P	85	39	0.13	0.16	0.03	0	0.06	0.59	0.04	0	0	0.62	0.23
Q	85	40	0.17	0.11	0.04	0	0.04	0.62	0.03	0	0	0.58	0.28
R	85	24	0.08	0.15	0.06	0	0.02	0.65	0.04	0	0	0.55	0.15
S	85	14	0	0.04	0	0	0	0.79	0.18	0	0	0.36	0.00
Y	85	13	0.17	0.04	0	0	0	0.53	0.26	0	0	0.65	0.29
Z	85	41	0.16	0.07	0.05	0	0	0.58	0.15	0	0	0.62	0.27

Chapter 3:

Habitat fragmentation and genetic variation: effects of climatically induced rising tree-line on an alpine butterfly

Introduction

Habitat fragmentation is of concern in conservation because of its potential to modify fundamental ecological processes at the population, community, and ecosystem levels (Burgess 1988; Fahrig & Merriam 1994; Saunders *et al.*1991). Habitat fragmentation may also affect microevolutionary processes such as gene flow and genetic drift, thereby altering levels of genetic variation (Templeton *et al.* 1990). Although the dynamics of genetic variation in subdivided populations can be complex (Varvio *et al.* 1986; Gilpin 1991; Chesser *et al.*1993; Wang 1997a), it is generally predicted that decreasing size and increasing isolation of habitat patches will lead to greater levels of genetic drift within and reduced gene flow among local populations (Templeton *et al.* 1990; Young *et al.* 1996). Over time, genetic variation within local populations will decline while differentiation among them increases (Templeton *et al.* 1990; Young *et al.* 1996). Loss of genetic variation can be highly detrimental because of inbreeding depression, lower individual and population viability, and impaired evolutionary potential (Ellstrand & Elam 1993). Although the importance of genetic factors in conservation has been questioned (Lande 1988; Caro & Laurenson 1994), an increasing body of empirical work suggests that genetic factors can affect the susceptibility of populations to endangerment and extinction (Frankham 1998; Saccheri *et al.* 1998; Madsen *et al.* 1999). Therefore, the potential impacts of habitat fragmentation on microevolutionary processes and genetic variation should not be ignored.

The Rocky Mountain Apollo, *Parnassius smintheus* Doubleday (1847), is a large and conspicuous butterfly of the mountain regions of western North America. It occurs in

alpine and subalpine meadows (Sperling & Kondla 1991), and is most commonly found above treeline. However, some populations in low-altitude, prairie habitats have also been observed (Bird *et al.* 1995). It has a wide distribution (Bird *et al.* 1995) and can be locally very abundant (Roland *et al.* 2000), but it appears to have fairly specific and restrictive habitat requirements (Fownes 1999) making its distribution quite patchy. Although this species is currently not considered threatened, its European congeners *Parnassius apollo* and *Parnassius mnemosyne* are of significant conservation interest (Descimon & Napolitano 1993; Witkowski & Adamski 1996; Meglécz *et al.* 1999; Kuras *et al.* 2000). Endangerment of these species has been linked to habitat loss and habitat fragmentation (Witkowski & Adamski 1996; Kuras *et al.* 2000).

Being typical of higher-altitude mountainous environments, much of the habitat of *P. smintheus* occurs in either protected areas or remote areas that are somewhat guarded from the more common anthropogenic agents of habitat fragmentation such as agriculture and urbanization. However, these habitats are threatened by a different agent that has little regard for park boundaries or rugged terrain: global climatic warming. Warming temperatures have already resulted in altitudinal shifts in alpine vegetation in this century (Grabherr *et al.* 1994), including increases in the elevational limit of tree-line (Kearney 1982; Taylor 1995; Woodward *et al.* 1995). This trend is only likely to accelerate with future climate change. Rising tree-line causes fragmentation of the high-altitude meadows that are inhabited by *P. smintheus* and a host of other alpine butterfly species.

Using mark-recapture methods we have found that movement of *P. smintheus*, though generally restricted, is particularly obstructed through forested areas (Roland *et al.* 2000). We have also shown that the degree of genetic differentiation among sample sites depends more strongly on the amount of intervening forest than on intervening meadow habitat (Keyghobadi *et al.* 1999). These results clearly indicate that habitat fragmentation by forestation will lead to greater genetic differentiation among local populations. Based on these findings, we have predicted that fragmentation of this species' habitat by encroaching

forests will also ultimately result in reduced genetic variation within local populations (Keyghobadi *et al.* 1999). Here, we test this prediction by looking at levels of genetic variation within local populations of *P. smintheus* and their relationship to patch characteristics such as size and connectivity. In particular, we test the following hypotheses about the effects of habitat fragmentation on microevolutionary processes:

1. Genetic drift is accentuated in smaller populations. Therefore, we expect that genetic variation within local populations will be positively correlated with population size and patch size.
2. Incoming gene flow counteracts the effects of drift by introducing genetic variation, and is greater in local populations that are less isolated. Therefore, genetic variation within local populations should be positively correlated with patch connectivity.
3. The type of habitat separating local populations (*i.e.* forested versus open meadow) is an important component of connectivity that affects levels of gene flow. Therefore, genetic variation should be more strongly correlated with a measure of patch connectivity that incorporates differential effects of forest and meadow habitat on rates of movement than with a measure that does not account for such effects.

Overall, this study represents an extension of our earlier investigations of the effects of landscape on population genetic structure in this species. We had previously considered the effects of landscape on genetic differentiation among local populations (Keyghobadi *et al.* 1999), and we now consider the effects of landscape characteristics on levels of genetic variation within local populations.

Materials and methods

Study area

Our study was conducted in a series of alpine meadows atop three ridges in the eastern edge of the Canadian Rocky Mountains in Kananaskis, Alberta: Lusk Ridge (51°00'N, 114°58'W), Jumpingpound Ridge (50°57'N, 114°53'W), and Cox Hill (50°58'N,

114°54'W). These meadows are dominated by grasses, sedges, and alpine wildflowers and are bordered by even-aged stands of lodgepole pine (*Pinus contorta*), subalpine fir (*Abies lasiocarpa*), and Engelmann spruce (*Picea engelmannii*). The meadows have been divided into 22 study sites (Fig. 3-1; Roland *et al.* 2000). Some sites are distinct patches of meadow surrounded by forest, while others are subsections of larger meadows that have previously been sub-divided for testing the differential effects of intervening forest and meadow habitat on butterfly movement using mark-recapture methods (Roland *et al.* 2000). All sites support *P. smintheus* and its sole hostplant, *Sedum lanceolatum*, in varying densities (Roland *et al.* 2000; Matter *et al.* in prep.). Also, within large meadows, the distributions of both *P. smintheus* and its hostplant are patchy. Because of time constraints, sites C and Z2 were not extensively sampled in the mark-recapture study or included in the analyses of those data (Roland *et al.* 2000). Samples for genetic analysis were also not collected from those sites, nor from d2, H, and N in which population sizes were too small to permit collection of sufficiently large samples.

Mark-recapture data and estimation of population size

Mark-recapture studies were conducted in most sites in 1995 and 1996 to investigate effects of landscape structure and population size on movement of individuals (Roland *et al.* 2000). Each site was visited four or five times over a five week period in each year. Adult butterflies were captured with hand nets and the lower, ventral portion of one hind wing was marked with a unique three-letter identifier using a fine-tipped permanent marker. Very few individuals were captured on more than two different sampling dates, but there were often many recaptures within a sampling date. Thus, we chose to use Craig's (1953) method to obtain estimates of population size within each site. This method assumes that the captures on a given day follow a Poisson distribution, from which the zero class (*i.e.*, the number of butterflies that were never caught) is estimated. This is added to the number of butterflies that were caught to obtain an estimate of the total number of butterflies present

on that sampling date. For each site, these values were averaged over all sampling dates, giving an estimate of the average number of adult butterflies present on a single day for each site. This estimate can be used as a relative index of population size. Other methods of analysis such as the Jolly-Seber method or Manly and Parr's method (Southwood 1978) could not be used because of the low number of multiple recaptures between sampling dates.

Measurement of landscape variables

Using 1993 black and white aerial photographs (1:40,000), the areas (ha) of all study sites were estimated. Three distance measures were also quantified for all pairs of sites: (1) total distance (km) which comprises (2) the distance through forest (km) and (3) the distance through meadow (km). All distances were measured between the centroids of butterfly observations within each site and along the ridges, since movements of *P. smintheus* are largely constrained along ridge-tops (Roland *et al.* 2000). The abundance of the hostplant, *S. lanceolatum*, in each site was estimated using transect counts in 1995, with additional censusing in 1998 (Matter *et al.* in prep.).

Measures of population size

For each study site we had four measures of population size that we could relate to genetic variation: (1) relative indices of population size calculated using Craig's (1953) method in 1995, (2) the same indices of population size for 1996, (3) area of the site or patch size (ha), and (4) abundance of the host plant in the site (total number of plants). Following convention (Frankham 1996), we examined the correlation between genetic variation and the logarithm of each measure of population size.

Index of patch connectivity

For each site, we measured its connectivity as

$$\text{Connectivity}_i = \sum_{j \neq i} \exp(-\alpha_f d_{fij} - \alpha_m d_{mij}) A_j^{(\xi_p - \xi_{em})}$$

where df_{ij} is the distance through forest between sites i and j, d_{mij} is the distance through meadow between sites i and j, and α_f and α_m are parameters that scale the rate of movement to distance for each of the two habitat types, respectively. A_j is the area of patch j. ξ_p is a parameter that scales population size to patch area and ξ_{em} scales the probability of emigration to patch area. The summation is over all other sites in the study area.

This is a slightly modified version of the patch connectivity defined by Hanski (1994) in the spatially explicit incidence function model of metapopulation dynamics. The measure defined by Hanski (1994) was meant to indicate the tendency for a patch to be colonized. By analogy, our measure is meant as an indicator of the total number of immigrants, and therefore of gene flow, into a site from all other possible sites in the study area. Our measure of connectivity assumes that immigration declines exponentially with distance at a rate determined by the α constants, with separate values accounting for differential effects of intervening forest and of meadow on movement. Also, the potential for a site to serve as a source of immigrants depends on its population size. This is assumed to be a function of the patch size, scaled to ξ_p to account for the fact that population size often scales non-linearly with patch area. ξ_{em} accounts for the possibility that the rate of emigration from a patch depends on its size. Although a detailed analysis of the mark-recapture data has found that emigration rate does not depend on patch area when other factors are controlled, it does depend on population size (Roland *et al.* 2000). Individuals are more likely to leave small populations than larger ones, perhaps reflecting selection of higher quality habitats (Roland *et al.* 2000). Thus, emigration rate may be inversely related to patch area depending on the relation between patch area and population size.

Estimates of α_f, α_m, and ξ_{em} were obtained using the Virtual Migration Model (VMM) (Hanski *et al.* 2000), which provides maximum-likelihood estimates of various parameters of migration and survival from mark-recapture data. We used the mark-recapture data from 1995, which fit the model better than did the data from 1996 (Matter *et al.* in prep.). We used a slightly modified version of the VMM that allows for estimation of more than one α value (Matter *et al.* in prep.). The estimated parameter values were: α_f= 5.81, α_m = 1.34, and ξ_{em} = 0.0. The value of $\xi_p$was estimated to be 0.84 by fitting a power function to the relationship between patch area and the average of the population size indices from 1995 and 1996. Population sizes fluctuate yearly and we expected that the average of the two years would be a slightly better indicator of long-term population size.

It should be noted that, although similar in form, our measure of connectivity is different from the measure defined in the Virtual Migration Model (Hanski *et al.* 2000). That measure indicates the tendency for a particular patch to send migrants out to all other patches in the study area, the opposite of what our measure indicates.

Sample collection , DNA extraction, and microsatellite genotyping

Samples for genetic analysis were collected from 17 of the 22 sites (Fig. 3-1) in 1995 and 1996 (Keyghobadi *et al.* 1999). Samples were either small (approx. 0.15 cm^2) wing clippings or whole adult butterflies. All samples were placed individually in glassine envelopes and stored at -80 °C. Genomic DNA was isolated from samples using the QIAamp™ tissue extraction kit (QIAGEN). For wing clippings, the entire sample was used. For whole butterflies, a sample of abdominal tissue (for males) or thoracic tissue (for females) was used. Each sample was typed at seven microsatellite loci (Ps50, Ps81, Ps85, Ps76, Ps163, Ps165, and Ps262) as described previously (Keyghobadi *et al.* 1999; Chapter 6), using one fluorescently labeled primer per locus (labeled with 6-FAM, HEX, or TET; Applied Biosystems). These samples have previously been typed at locus Ps162 as well (Keyghobadi *et al.* 1999). However, this locus was dropped from the analysis as

we have found it to amplify multiple bands in some individuals sampled from a different area (Banff National Park). Products of polymerase chain reaction (PCR) amplification were electrophoresed and detected using an Applied Biosystems 373A Automated Sequencer and analyzed using GENESCAN and GENOTYPER software (Applied Biosystems).

Statistical analysis

For the samples analyzed here, conformity of genotypes at the seven microsatellite loci to Hardy-Weinberg proportions have previously been tested (Keyghobadi *et al.*1999; Chapter 6). Because we have found excess homozygosity and evidence of null alleles at all loci (Keyghobadi *et al.* 1999; Chapter 6), for each locus at each site the frequency of the null allele was estimated, and the frequencies of all other alleles were simultaneously re-estimated, using the estimation-maximization (EM) algorithm (Ceppellini *et al.* 1955; Yasuda & Kimura 1968; Long *et al.* 1995), as described in Keyghobadi *et al.* (1999) [calculator can be found at <http://www.biology.ualberta.ca/jbrzusto/nullele>]. These allele frequencies were used to calculate unbiased estimates of expected heterozygosity for each locus within each site, as $(1-\Sigma p_i^2)(2N/2N-1)$, where p_i is the estimated frequency of allele i, and N is the number of individuals sampled (Nei & Roychoudhury 1974). Within each site, expected heterozygosities were averaged over loci to give an overall indicator of genetic variability, the average expected heterozygosity (H_e). In choosing an indicator of genetic variability within sites, we had several other options: number of alleles, observed heterozygosity, or the probability of identity (Paetkau & Strobeck 1994). We did not use the number of alleles because it tends to be strongly influenced by sample size. Observed heterozygosity is inversely related to the null allele frequency and therefore is not a reliable indicator of within site variability, particularly if null allele frequencies vary among sites. The probability of identity can be calculated from the allele frequencies estimated using the EM algorithm, as we have done for H_e. However, the probability of identity represents

essentially a different expression of the same data used to calculate H_e. Given that H_e is more commonly used in the literature and that most theoretical developments consider responses of heterozygosity to various evolutionary forces (Hedrick & Gilpin 1997), we chose to limit our analysis to H_e.

The significance of correlations between H_e and indices of population size or patch connectivity were tested using the nonparametric Kendall's rank correlation (Sokal & Rohlf 1995). In all cases we present both Kendall's τ and the r^2 value from standard parametric correlation.

Results

Effects of intervening landscape on genetic differentiation among sites

The effects of intervening landscape on levels of genetic differentiation among these same sites have previously been investigated using four microsatellite loci (Keyghobadi *et al.* 1999). Here, we typed the same set of samples at four additional loci and dropped one of the previously used loci (Ps162). Re-analysis of the effects of landscape on population genetic structure using these new molecular data did not alter the general results or conclusions presented before (Keyghobadi *et al.* 1999). In particular, there was still a positive correlation between Nei's standard genetic distance (Nei 1972) and geographic distance (Mantel test: $r = 0.54$, $P = 0.0003$). Also, partial Mantel tests (Smouse *et al.* 1986) showed that the amount of forest separating sites explained more variation in genetic distance than did the amount of meadow separating sites: the partial correlation between distance through forest and Nei's standard genetic distance, controlling for distance through meadow, was significant ($r = 0.31$, $P = 0.02$), while the partial correlation between distance through meadow and Nei's standard genetic distance, controlling for distance through forest, was not ($r = 0.14$, $P = 0.06$).

Levels of genetic variation within sites

Genetic variation within sites was generally high but variable, with the average number of alleles at seven microsatellite loci ranging from 6.7 to 11.3 and the average expected heterozygosity (H_e) ranging from 0.643 to 0.766 (Table 3-1). Sample size was positively correlated with the average number of alleles ($r^2 = 0.80$, $P = 0.0001$), but not with H_e ($r^2 = 0.22$; $\tau = 0.01$, $P = 0.93$), confirming our decision to use H_e as a measure of within-site genetic variability in our analyses.

Genetic variation and population size

None of the measures of population size were correlated with H_e (Fig. 3-2). Some of our sampling sites were carved out of larger meadows. Given that both rates of movement and genetic similarity among sites are reduced much more strongly by intervening forest than meadow (Keyghobadi *et al.* 1999; Roland *et al.* 2000), it may be more appropriate to consider each patch of uninterrupted meadow as a single population. We re-analyzed the relationships between the various measures of population size and H_e with all sites that were situated in a single, uninterrupted patch of meadow considered as part of one population. That is, for all sites in a single meadow allele frequencies were averaged to calculate H_e and the measures of population size were summed over the sites. There was still no significant correlation between H_e and any of the measures of population size (Fig. 3-3).

There was also no correlation between He and the harmonic mean of the population size indices from 1995 and 1996 (ln transformed), whether sites within a meadow were considered separately ($r^2 = 0.01$; $\tau = -0.04$, $P = 0.80$) or as part of a single population ($r^2 = 0.04$; $\tau = 0.09$, $P = 0.70$).

Genetic variation and patch connectivity

There was a significant positive correlation between H_e and patch connectivity (Fig. 3-4: $r^2 = 0.29$; $\tau = 0.37$, $P = 0.04$). Our study area is dominated by the large, forested valley separating Lusk Ridge (sites C, D1, d2, and E) from Jumpingpound Ridge and Cox Hill (all other sites), and this may have an inordinate influence on our results. However, removal of the Lusk Ridge sites from the analysis, including the calculations of patch connectivity of other sites, had little effect on the correlation between H_e and patch connectivity ($r^2 = 0.30$; $\tau = 0.39$, $P = 0.04$). This is not surprising given that connectivity is largely determined by the immediately neighbouring patches, and sites located across the large valley have very little effect on each others' connectivity.

When all sites located within a single, uninterrupted meadow were considered as one population, and measures of H_e and patch connectivity were adjusted accordingly, the correlation between H_e and connectivity decreased appreciably ($r^2 = 0.17$; $\tau = 0.18$, $P = 0.41$).

In our calculations of patch connectivity we used patch area (A_j), scaled to an exponent (ξp), as an indicator of population size. The value of the exponent (ξp) was estimated by fitting the average of the population size indices from 1995 and 1996 to patch area. Surprisingly, replacing the term $A_j^{\xi p}$ with the actual average of the population size indices from 1995 and 1996 reduced the correlation between H_e and connectivity ($r^2 = 0.14$; $\tau = 0.24$, $P = 0.19$). In re-calculating connectivity using the population size indices, we had no population size estimates for site Z2 and we did not include that patch in the measures of connectivity for the other patches. This most strongly affected the measure of connectivity for the adjacent site Z1, resulting in an underestimate, but had little effect on measures of connectivity for all other sites. Therefore, we excluded site Z1 from the correlation analysis. The reduction in the correlation was not due to the exclusion of this data point however, because when connectivity was calculated using patch areas [$A_j^{(\xi p)}$], excluding Z2 from the calculations of patch connectivity and excluding Z1 from the

correlation analysis still resulted in a significant correlation between H_e and patch connectivity ($r^2 = 0.30$; $\tau = 0.38$, $P = 0.04$).

Another surprising result was that not accounting for differential effects of intervening forest and meadow habitat in the measures of patch connectivity (*i.e.* only using total distance between patches) led to an increase in the observed correlation between H_e and connectivity. Thus, if we assumed that movement across both habitat types was equivalent to movement over open meadow and calculated connectivity as

$$\text{Connectivity}_i = \sum_{j \neq i} \exp(-\alpha_m d_{ij}) A_j^{(\zeta_j - \zeta_m)}$$

where d_{ij} is the total distance between patches i and j and $\alpha_m = 1.34$ is the parameter scaling movement to distance over meadow , then the correlation between He and connectivity increased (Fig. 3-5; $r^2 = 0.39$; $\tau = 0.47$, $P = 0.008$).

Correlation of heterozygosity with patch areas and connectivities from 1952

Time lags in the response of genetic variation to altered rates of microevolutionary forces following fragmentation can lead to low correlation between observed patterns of genetic variation and contemporary patterns of habitat fragmentation (Young *et al.* 1993). Because of a large fire that occurred on Jumpingpound Ridge in 1938, the open, non-forested habitat on this ridge was much more extensive in the past. Since that time, re-establishment of forests and fragmentation of the ridge-top meadows has occurred. We used landscape data (*i.e.* patch areas and inter-patch distances) from 1952 aerial photographs (1: 40,000) to examine the extent to which past landscape was correlated with current levels of genetic variation (Fig. 3-6). There was no significant correlation between patch area in 1952 (ln transformed) and H_e ($r^2 = 0.01$; $\tau = 0.03$, $P = 0.87$). However, patch connectivity in 1952 was better correlated with current He ($r^2 = 0.34$; $\tau = 0.43$, $P = 0.02$) than was patch connectivity in 1993 (Fig. 3-7).

Discussion

Hypothesis 1: Genetic variation and population size

We did not find a significant correlation between average heterozygosity (H_e) and any measure of population size (Fig. 3-2, Fig. 3-3). Population genetic theory predicts a positive correlation between genetic variability and population size, because genetic variation in finite populations decays through the process of drift at a rate determined by the effective population size (Wright 1931). Effective population size, in turn, is considered to be some proportion of the total population size (Frankham 1995). In general, genetic variation is positively correlated with population size in natural populations (Frankham 1996), however there are many exceptions (e.g., Desender *et al.* 1998; da Silva & Granadeiro 1999; Madsen *et al.* 2000; Schmidt & Jensen 2000; Shapcott 2000).

One may not observe a correlation between population size and genetic variation if effective population sizes are all large enough that drift is negligible. For example, Soulé (1980) identified a threshold size of approximately 250-2000 above which no correlation between genetic variability and population size was observed in two species of lizards. Within each of our study sites, estimates of the number of adult butterflies present per day, averaged over all sampling dates, ranged from 0 to 230 and were generally higher in 1996 than in 1995 (Table 3-1). Since the adult flight season lasts about 8 weeks and most adults live for approximately 10 days (Roland *et al.* 2000), then we can estimate that the total adult population sizes within each of these sites ranged from approximately 0 to 1300 individuals, and averaged 270 (excluding estimates of zero population size) over the two years. On average, effective sizes of natural populations have been estimated to be approximately one tenth of the census population size in a variety of species (Frankham 1995). Thus, we estimate that the effective population sizes in our study sites averaged 27, with a maximum value of 130. Even if effective population size is as high as one half of census population size, we are left with an average estimate of effective population size of 135 with a maximum value of 650. Though these estimates are very rough, they do

suggest that effective population sizes in most of our sites are of the order of magnitude that drift should play a significant role in determining levels of genetic variation (Soulé 1980).

A commonly invoked explanation for lack a positive correlation between current population sizes and genetic variability is the influence of historical events, particularly population bottlenecks, on levels of genetic variation (e.g., Ellstrand & Elam 1993; da Silva & Granadeiro 1999; Schmidt & Jensen 2000). When population size is not constant over time, the effective population size is a function of the harmonic mean of the census population sizes over the years and is heavily influenced by the minimum population size (Wright 1931; Wright 1938; Lande & Barrowclough 1987). Thus, historical population bottlenecks can have a significant impact on currently observed levels of genetic variation. For example, Bodkin *et al.* (1999) found that in sea otter populations that had experienced bottlenecks, genetic diversity was correlated with minimum population size and the number of years at minimum population size instead of contemporary population size. Population sizes of *P. smintheus* do fluctuate yearly (Roland *et al.* 2000; pers. obs.), as is typical of many temperate butterfly species (Ehrlich 1965; Watt *et al.* 1977; Pollard & Yates 1993). On occasion, population numbers could drop appreciably, for example as a result of drought or other extreme weather events (e.g.Watt *et al.* 1977). Thus, levels of genetic variation might be better correlated with the harmonic mean population size over a number of years or the minimum population size rather than the population size in one or a few recent years. In addition, the minimum population sizes might be uncorrelated with patch characteristics such as size and hostplant abundance if the minimum population sizes occur as a result of extreme weather or other 'catastrophic' events. Monitoring population sizes in the study area over the years will shed light on the extent of annual fluctuations and on whether this is a likely explanation for the lack of correlation between genetic variation and current local population size.

Finally, the lack of correlation between H_e and measures of population size may simply reflect the fact that, in a subdivided population with ongoing gene flow,

heterozygosity is not simply a function of local effective population size but also depends on the rate of migration among local demes (Hartl & Clark 1989).

Hypothesis 2: Genetic variation and patch connectivity

We found a significant, positive correlation between H_e and patch connectivity (Fig. 3-4). Presumably, higher connectivity allows for a higher rate of incoming gene flow which counters the effects of drift and has a positive effect on levels of genetic variation. This result confirms our concerns about the potential impact of increasing patch isolation to reduce within-patch genetic variation in this species. A number of studies have found a significant, positive correlation between various measures of patch connectivity and genetic variation (Soulé 1972; Saura *et al.* 1973; Descimon & Napolitano 1993; Schmitt *et al.* 1995; Hänfling & Brandl 1998; Shapcott 2000; Vucetich *et al.* 2001), though exceptions also exist (Young *et al.* 1999; Jäggi *et al.* 2000).

Replacing patch areas by the average indices of population size in the measures of patch connectivity reduced the correlation between patch connectivity and H_e. This is surprising because patch area is a surrogate of population size and is meant to measure the potential of a site to act as a source of migrants. However, given yearly fluctuations in population size, patch area may be a better long-term estimate of the potential number of emigrants from a patch than is the population size in one or two particular years. Alternatively, this result suggests that the indices of population size may be relatively poor estimators of the true population size.

Hypothesis 3: Importance of intervening landscape as a component of patch connectivity

We found that not accounting for differences in rates of movement through forest versus meadow habitats in our measure of patch connectivity increased the correlation between connectivity and H_e (Fig. 3-5). This result was very puzzling given that both levels of genetic differentiation among sites and rates of movement are clearly related to the relative amounts of intervening forest versus meadow habitat (Keyghobadi *et al.* 1999; Roland *et*

al. 2000). We would expect a measure of patch connectivity that accounts for the different effects of these landscape elements to be more accurate and therefore better correlated with the genetic variability within patches. However, we have found the apparently paradoxical result that differentiation among sites depends on the nature of intervening landscape, while variation within sites does not. We speculate that this is due to different rates of response of among-site genetic differentiation versus within-site genetic variation to habitat fragmentation. A large fire occurred on Jumpingpound Ridge in 1938 after which the open, non-forested habitat on the ridge was quite extensive (Fig. 3-6). Since that time, the encroachment of forest into formerly open areas has progressively reduced the size and connectivity of the meadows. We propose that as forests have advanced into the meadows causing their fragmentation, rates of gene flow among local populations have declined; however, genetic differentiation among sites has responded more rapidly to the fragmentation than has the loss of variation within sites. Thus, although genetic distances among sites are correlated with the current state of the landscape, levels of variation within sites are not. This hypothesis is supported by theoretical examinations of the dynamics of genetic variation in subdivided populations which suggest that differentiation among sites and variation within sites can approach equilibrium values at very different rates. For example, Wang (1997a; b) found that with subdivision of a previously panmictic population, the gene correlations within individuals, within subpopulations, and among subpopulations could initially change at very different rates. Also Varvio *et al.* (1986) found that under a wide range of conditions, the gene diversity within local populations (H_S, which is equivalent to H_e) reached equilibrium values much more slowly than did the genetic differentiation among local populations (G_{st}). Our hypothesis is further supported by the observation that H_e is better correlated with patch connectivity 40 years previous than it is with patch connectivity today (Fig. 3-4; Fig. 3-7).

Populations in nature may not be at equilibrium with respect to the action of microevolutionary forces (Varvio *et al.* 1986; Whitlock & McCauley 1999). With reference

to studying the effects of habitat fragmentation on genetic variation, it has been noted that observed levels and patterns of genetic variation may not reflect the current degree of fragmentation because of time lags to equilibrium (Young *et al.* 1993). Our results underline a related, but different point: in subdivided populations, different aspects of genetic structure (*i.e.* differentiation among versus variation within local populations) may approach equilibrium, and therefore respond to habitat fragmentation, at different rates.

Sampling units

When testing for relationships between genetic variation and population or patch characteristics such as size and isolation, one assumes that the sampled units are discrete demographic units that experience drift at a rate determined by the local effective population size and that also experience some level of gene flow from other demographic units. Deviations from this assumption or sampling at an inappropriate scale can obscure any relationships between genetic variation and population or patch characteristics. The lack of correlation between measures of population size and H_e (Fig. 3-2) may be due to such errors in sampling.

Given the small spatial scale of our study, it is unlikely that we have erroneously combined independent demographic units in a single sample. However, it is possible that we have split what are essentially single demographic units into two or more separate samples. In that case, we will have underestimated population sizes and incorrectly estimated patch connectivities. Since movement and gene flow among sites are reduced much more strongly by intervening forest than by meadow (Keyghobadi *et al.* 1999; Roland *et al.* 2000), it may be more appropriate to consider each patch of uninterrupted meadow as a single population. However, when we did this, there was little effect on the relationship between H_e and measures of population size (Fig. 3-3), and the correlation between H_e and connectivity decreased considerably. This suggests that our division of large meadows into two or more sampling units was probably a better approximation to the

true population structure in the study area. This is consistent with our observations that movement in this species, even over open meadows, is quite restricted. For example, in the mark-recapture study, the average distance moved by adult butterflies was only approximately 150m (Roland *et al.* 2000). In contrast, the average distance between centroids of sites located in the same large meadow was 450m. Strong population structure over a very small spatial scale and in the absence of obvious geographic barriers has also been observed in the butterfly *Euphydryas editha* (Ehrlich 1961).

It is not immediately obvious what other sampling scheme may be an even better approximation to the true population structure in this series of habitat patches. When local demographic units are connected by ongoing gene flow it may be very difficult, if not impossible, to delimit boundaries among them and to define local population sizes. This source of error cannot be ruled out as a potential explanation for the lack of correlation between H_e and measures of population size. Measures of patch connectivity are probably less sensitive to this problem than measures of local population size because connectivity incorporates information on the sizes and distances of the surrounding local demographic units.

Selection and mutation

In the interpretation of our results, we have assumed that differences in H_e among sites are determined by differences in the levels of genetic drift and gene flow, and we have ignored possible differences in selection or mutation rates. Although there are exceptions (e.g., Caskey *et al.* 1992), variation at microsatellite loci is generally considered to be selectively neutral (Jarne & Lagoda 1996). Therefore, the assumption that there is negligible among-site variation in selection at these markers seems reasonable.

There is also no reason to suspect that the mutation rates at the microsatellite loci we have used vary among sample sites. It has been suggested that mutation rates at microsatellite loci are positively related to levels of heterozygosity because mutations are

more likely to occur in individuals with a greater size difference between their alleles (Amos *et al.* 1996). This would result in a positive feedback between heterozygosity and mutation rate which would exacerbate initial differences in heterozygosity. However, a relationship between heterozygosity and mutation rate has been questioned (Ellegren 2000) and is incompatible with the generally accepted model of microsatellite mutation by replication slippage. Even if such a relationship does exist, we have no reason to suspect that it would lead to spurious correlations between heterozygosity and patch characteristics, only that it may amplify any such correlations that already exist.

Immigration from outside the study area

We have ignored the possibility of immigration or gene flow from outside the study area. The closest potential sources of immigrants from outside the study area are a population on Moose Mtn. approximately 5 km to the east and a population near Mt. Baldy approximately 6.25 km to the west. These sites are far enough that including them in the measures of patch connectivity for sites in our study area would have had little effect on the calculated values. Therefore, we do not believe that our measures of connectivity or our results in general are affected by the possibility of immigration from outside the study area.

Long-term implications of habitat fragmentation for P. smintheus

Overall, our results clearly indicate that increasing isolation of local populations, in terms of greater distances among the populations and reduced sizes of surrounding populations, does have a negative impact on local levels of genetic variation. However, the influence of changing relative amounts of intervening forest versus meadow habitat in the landscape is equivocal. Intervening forests increase genetic differentiation among sites, suggesting that increased forest cover in the landscape reduces gene flow and therefore should also reduce local genetic variation (Keyghobadi *et al.* 1999). But, there appears to be little relationship between within-patch H_e and the current distribution of forest and meadow habitats in the

landscape. We have suggested that levels of variation within sites, in comparison to levels of differentiation among sites, have responded slowly to the habitat fragmentation that has occurred in our study area following a fire in the 1930s that cleared much of the forest. Therefore, we maintain that habitat fragmentation by forest encroachment has negative consequences for levels of genetic variation within local populations of *P. smintheus*, though the response may be slow. Habitat fragmentation by rising tree-line will, in the long-term, be detrimental to this alpine species and provides an example of how global climate change can threaten even those species living in protected or remote areas (Peters & Darling 1985).

Because reduced gene flow is predicted to lead to lower variation within but also greater differentiation among local populations, regional levels of genetic diversity may not change or may even increase in response to habitat fragmentation. Thus, it has been predicted that in the face of habitat fragmentation it is possible for much of the original genetic variation in a given area to be preserved (Gilpin 1987; Templeton *et al.* 1990). However, reduced genetic variation is considered detrimental to populations because of the effects of inbreeding in the short term (Soulé 1980) and a reduction in evolutionary potential in the long term (Franklin 1980). In the absence of management intervention, among-site variation will be of little value if the processes of inbreeding and adaptation occur at the level of local populations, which is presumably the case when local populations are fairly isolated.

Finally, as a caveat, we should note that we have used variation at presumably neutral genetic markers to test hypotheses about how the processes of drift and gene flow will be affected changes in landscape structure. The actual response of adaptively significant variation to changes in these evolutionary processes will also depend on the magnitude and direction of selective pressures (Lande & Barrowclough 1987) and therefore, may ultimately differ from the response of neutral variation.

Literature Cited

Amos, W., S. J. Sawcer, R. W. Feakes, and D. C. Rubinsztein. 1996. Microsatellites show mutational bias and heterozygote instability. *Nature Genetics* **13**:390-391.

Bird, C. D., G. J. Hilchie, N. G. Kondla, E. M. Pike, and F. A. H. Sperling. 1995. Alberta Butterflies. Provincial Museum of Alberta, Edmonton.

Bodkin, J. L., B. E. Ballachey, M. A. Cronin, and K. T. Scribner. 1999. Population demographics and genetic diversity in remnant and translocated populations of sea otters. *Conservation Biology* **13**:1378-1385.

Burgess, R. L. 1988. Community organization: effects of landscape fragmentation. *Canadian Journal of Botany* **66**:2687-2690.

Caro, T. M., and M. K. Laurenson. 1994. Ecological and genetic factors in conservation: a cautionary tale. *Science* **263**:485-486.

Caskey, C. T., A. Pizzuti, Y.-H. Fu, R. G. Fenwick Jr., and D. L. Nelson. 1992. Triplet repeat mutations in human disease. *Science* **256**:784-789.

Ceppellini, R., M. Siniscalco, and C. A. B. Smith. 1955. The estimation of gene frequencies in a random-mating population. *Annals of Human Genetics* **20**: 97-115.

Chesser, R. K., O. E. Rhodes Jr., D. W. Sugg, and A. Schnabel. 1993. Effective sizes for subdivided populations. *Genetics* **135**:1221-1232.

Craig, C. C. 1953. On the utilization of marked specimens in estimating populations of flying insects. *Biometrika* **40**:170-176.

da Silva, M. C., and J. P. Granadeiro. 1999. Genetic variability and isolation of Cory's shearwater colonies in the northeast Atlantic. *The Condor* **101**:174-179.

Descimon, H., and M. Napolitano. 1993. Enzyme polymorphism, wing pattern variability, and geographical isolation in an endangered butterfly species. *Biological Conservation* **66**:117-123.

Desender, K., T. Backeljau, K. Delahaye, and L. DeMeester. 1998. Age and size of European saltmarshes and the population genetic consequences for ground beetles. *Oecologia* **114**:503-513.

Ehrlich, P. R. 1961. Intrinsic barriers to dispersal in checkerspot butterfly. *Science* **134**:108-109.

Ehrlich, P. R. 1965. The population biology of the butterfly, *Euphydryas editha*. II. The structure of the Jasper Ridge colony. *Evolution* **19**:327-336.

Ellegren H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**: 551-558.

Ellstrand, N. C., and D. R. Elam. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* **24**:217-242.

Fahrig, L., and G. Merriam. 1994. Conservation of fragmented populations. *Conservation Biology* **8**:50-59.

Fownes, S. L. 1999. Meadow suitability and quality for the alpine butterfly, *Parnassius smintheus*, in the east slopes of the Canadian Rockies. M.Sc Thesis. University of Alberta.

Frankham, R. 1995. Effective population size / adult population size ratios in wildlife: a review. *Genetical Research* **66**:95-107.

Frankham, R. 1996. Relationship of genetic variation to population size in wildlife. *Conservation Biology* **10**:1500-1508.

Frankham, R. 1998. Inbreeding and extinction: island populations. *Conservation Biology* **12**:665-675.

Franklin, I. R. 1980. Evolutionary change in small populations, pp. 135-149. In M. E. Soulé and B. A. Wilcox (eds.), Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer Associates Inc., Sunderland, Massachusetts.

Gilpin, M. 1991. The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society* **42**:165-175.

Gilpin, M. E. 1987. Spatial structure and population vulnerability, pp. 125-139. In M. E. Soulé (ed.), Viable Populations for Conservation. Cambridge University Press, Cambridge.

Grabherr, G., M. Gottfried, and H. Pauli. 1994. Climate effects on mountain plants. *Nature* **369**:448.

Hänfling, B., and R. Brandl. 1998. Genetic variability, population size and isolation of distinct populations in the freshwater fish *Cottus gobio* L. *Molecular Ecology* **7**:1625-1632.

Hanski, I. 1994. A practical model of metapopulation dynamics. *Journal of Animal Ecology* **63**:151-162.

Hanski, I., J. Alho, and A. Moilanen. 2000. Estimating the parameters of survival and migration of individuals in metapopulations. *Ecology* **81**:239-251.

Hartl D. L., and A. G. Clark. 1989. Principles of Population Genetics. 2nd edition. Sinauer Associates Inc., Sunderland, Mass.

Hedrick, P. W., and M. E. Gilpin. 1997. Genetic effective size of a metapopulation, pp. 165-181. In I. A. Hanski and M. E. Gilpin (eds.), Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, San Diego.

Jäggi, C., T. Wirth, and B. Baur. 2000. Genetic variability in subpopulations of the asp viper (*Vipera aspis*) in the Swiss Jura mountains: implications for a conservation strategy. *Biological Conservation* **94**:69-77.

Jarne, P. and. P. J. L. Lagoda. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**:424-429.

Kearney, M. S. 1982. Recent seedling establishment at timberline in Jasper National Park, Alta. *Canadian Journal of Botany* **60**:2283-2287.

Keyghobadi, N., J. Roland, and C. Strobeck. 1999. Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* **8**: 1481-1495. **Chapter 2 of this volume.**

Kuras, T., J. Benes, A. Celechovsky, V. Vrabec, and M. Konvicka. 2000. *Parnassius mnemosyne* (Lepidoptera: Papilionidae) in North Moravia: review of present and past distribution, proposal for conservation. *Klapalekiana* **36**:93-112.

Lande, R., and G. F. Barrowclough. 1987. Effective population size, genetic variation, and their use in population management, pp. 87-123. In M. E. Soulé (ed.), Viable Populations for Conservation. Cambridge University Press, Cambridge.

Lande, R. 1988. Genetics and demography in biological conservation. *Science* **241**:1455-1460.

Long, J. C., R. C. Williams, and M. Urbanek. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. *American Journal of Human Genetics* **56**: 799-810.

Madsen, T., R. Shine, M. Olsson, and H. Wittzell. 1999. Restoration of an inbred adder population. *Nature* **402**:34-35.

Madsen, T., M. Olsson, H. Wittzell, B. Stille, A. Gullberg, R. Shine, S. Andersson, and H. Tegelström. 2000. Population size and genetic diversity in sand lizards (*Lacerta agilis*) and adders (*Vipera berus*). *Biological Conservation* **94**:257-262.

Matter, S. F., J. Roland, A. Moilanen, I. Hanski, and K. J. Sabourin. In Prep. The migration and survival of individuals: effects of habitat matrix type.

Meglécz, E., G. Nève, K. Pecsenye, and Z. Varga. 1999. Genetic variations in space and time in *Parnassius mnemosyne* (L.) (Lepidoptera) populations in north-east Hungary: implications for conservation. *Biological Conservation* **89**:251-259.

Nei, M. 1972. Genetic distance between populations. *American Naturalist* **106**: 283-292.

Nei, M., and A. K. Roychoudhury. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* **76**: 379-390.

Paetkau, D. and. C. Strobeck. 1994. Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* **3**:489-495.

Peters, R. L. and J. D. S. Darling. 1985. The greenhouse effect and nature reserves. *Bioscience* **35**: 707-717.

Pollard, E., and T. J. Yates. 1993. Monitoring butterflies for ecology and conservation. Chapman and Hall, London.

Roland, J., N. Keyghobadi, and S. Fownes. 2000. Alpine *Parnassius* butterfly dispersal: effects of landscape and population size. *Ecology* **81**: 1642-1653.

Saccheri, I., M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius, and I. Hanski. 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**:491-494.

Saunders, D. A., R. J. Hobbs, and C. R. Margules. 1991. Biological consequences of ecosystem fragmentation: a review. *Conservation Biology* **5**:18-32.

Saura, A., O. Halkka, and J. Lokki. 1973. Enzyme gene heterozygosity in small island populations of *Philaenus spumarius* (L.) (Homoptera). *Genetica* **44**:459-473.

Schmidt, K., and K. Jensen. 2000. Genetic structure and AFLP variation of remnant populations in the rare plant *Pedicularis palustris* (Scrophulariaceae) and its relation to population size and reproductive components. *American Journal of Botany* **87**:678-689.

Schmitt, L. H., D. J. Kitchener, and R. A. How. 1995. A genetic perspective of mammalian variation and evolution in the Indonesian archipelago: biogeographic correlates in the fruit bat genus *Cynopterus*. *Evolution* **49**:399-412.

Shapcott, A. 2000. Conservation and genetics in the fragmented monsoon rainforest in the Northern Territory, Australia: a case study of three frugivore-dispersed species. *Australian Journal of Botany* **48**:397-407.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**:627-632.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. 3rd edn. W. H. Freeman and Company, New York.

Soulé, M. 1972. Phenetics of natural populations. III. Variation in insular populations of a lizard. *American Naturalist* **106**:429-446.

Soulé, M. E. 1980. Thresholds for survival: maintaining fitness and evolutionary potential, pp. 151-169. In M. E. Soulé and B. A. Wilcox (eds.), Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer Associates, Inc., Sunderland, Massachusetts.

Southwood, T. R. E. 1978. Ecological Methods: With Particular Reference to the Study of Insect Populations. Chapman and Hall, London.

Sperling, F. A. H., and N. G. Kondla. 1991. Alberta Swallowtails and Parnassians: natural history, keys, and distribution. *Blue Jay* **49**:183-192.

Taylor, A. H. 1995. Forest expansion and climate change in the mountain hemlock (*Tsuga mertensiana*) zone, Lassen Volcanic National Park, California, USA. *Arctic and Alpine Research* **27**:207-216.

Templeton, A. R., K. Shaw, E. Routman, and S. K. Davis. 1990. The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Gardens* **77**:13-27.

Varvio, S., R. Chakraborty, and M. Nei. 1986. Genetic variation in subdivided populations and conservation genetics. *Heredity* **57**:189-198.

Vucetich, L. M., J. A. Vucetich, C. P. Joshi, T. A. Waite, and R. O. Peterson. 2001. Genetic (RAPD) diversity in *Peromyscus maniculatus* populations in a naturally fragmented landscape. *Molecular Ecology* **10**:35-40.

Wang, J. 1997a. Effective size and F-statistics of subdivided populations. I. Monoecious species with partial selfing. *Genetics* **146**:1453-1463.

Wang, J. 1997b. Effective size and F-statistics of subdivided populations. II. Dioecious species. *Genetics* **146**:1465-1474.

Watt, W. B., F. S. Chew, L. R. G. Snyder, A. G. Watt, and D. E. Rothschild. 1977. Populaton structure of Pierid butterflies I: Numbers and movements of some montane *Colias* species. *Oecologia* **27**:1-22.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. *Heredity* **82**:117-125.

Witkowski, Z., and P. Adamski. 1996. Decline and rehabilitation of the apollo butterfly *Parnassius apollo* (Linnaeus, 1758) in the Pieniny National Park (Polish Carpathians), pp. 7-14. In J. Settele C. Margules, P. Poschlod, and K. Henle (eds.), Species Survival in Fragmented Landscapes. Khuwer Academic Publishers, Dordrecht.

Woodward, A., E. G. Schreiner, and D. G. Silsbee. 1995. Climate, geography, and tree establishment in subalpine meadows of the Olympic Mountains, Washington, U.S.A. *Arctic and Alpine Research* **27**:217-225.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97-159.

Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* **87**:430-431.

Yasuda, N., and M. Kimura. 1968. A gene-counting method of maximum likelihood for estimating gene frequencies in ABO and ABO-like systems. *Annals of Human Genetics* **31**: 409-420.

Young, A., T. Boyle, and T. Brown. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution* **11**:413-418.

Young, A. G., H. G. Merriam, and S. I. Warwick. 1993. The effects of forest fragmentation on genetic variation in *Acer saccharum* Marsh. (sugar maple) populations. *Heredity* **71**:277-289.

Young, A. G., A. H. D. Brown, and F. A. Zich. 1999. Genetic structure of fragmented populations of the endangered daisy *Rutidosis leptorrhynchoides*. *Conservation Biology* **13**:256-265.

Table 3-1. Genetic variation, measures of population size, and measures of connectivity for study sites. Sample size refers to the number of individuals genotyped at seven microsatellite loci. Mean no. of alleles refers to the observed number of alleles (including null allele), averaged over loci. Avg. H_e is the unbiased estimate of heterozygosity, averaged over loci. Population indices for 1995 and 1996 were calculated using Craig's (1953) method.

Site	Sample size	Mean no. alleles	Avg. H_e	Patch area (ha)	Host plant abundance (no. of plants)	Population index (1995)	Population index (1996)	Patch connectivity
C	-	-	-	4.1	6417	10.0	-	6.05
D1	43	9.3	0.657	8.7	8700	38.0	120.7	4.26
d2	-	-	-	1.1	1133	1.2	0.0	6.75
E	40	9.7	0.708	9.2	22264	58.3	124.0	2.04
F	41	10.9	0.743	3.0	23295	31.6	65.5	2.68
G1	40	11.3	0.698	8.5	45985	60.0	84.1	4.23
g2	40	10.6	0.754	2.5	22887	42.4	133.0	6.95
H	-	-	-	3.4	2482	4.0	1.0	5.19
I	21	8.7	0.766	4.1	29766	1.0	65.0	13.27
J	31	10.1	0.732	26.3	38121	3.0	58.7	11.64
K	40	10.9	0.764	8.0	60400	18.3	40.1	16.18
L	40	10.0	0.715	18.5	52545	9.0	116.0	13.68
M	38	10.9	0.724	25.6	54528	26.5	230.0	8.91
N	-	-	-	1.3	3361	0.0	22.5	5.90
O	12	6.7	0.674	2.3	3335	4.3	13.3	3.70
P	39	11.0	0.751	7.7	6160	10.0	76.9	5.26
Q	40	10.6	0.727	10.2	10251	6.6	102.7	4.17
R	24	9.3	0.734	1.3	4030	6.9	36.0	2.52
S	14	6.9	0.643	15.1	8532	1.0	13.6	0.29
Y	13	7.4	0.712	0.8	5192	5.8	0.0	1.01
Z1	41	10.1	0.722	10.1	80199	58.5	121.3	3.51
Z2	-	-	-	56.5	-	-	-	0.82

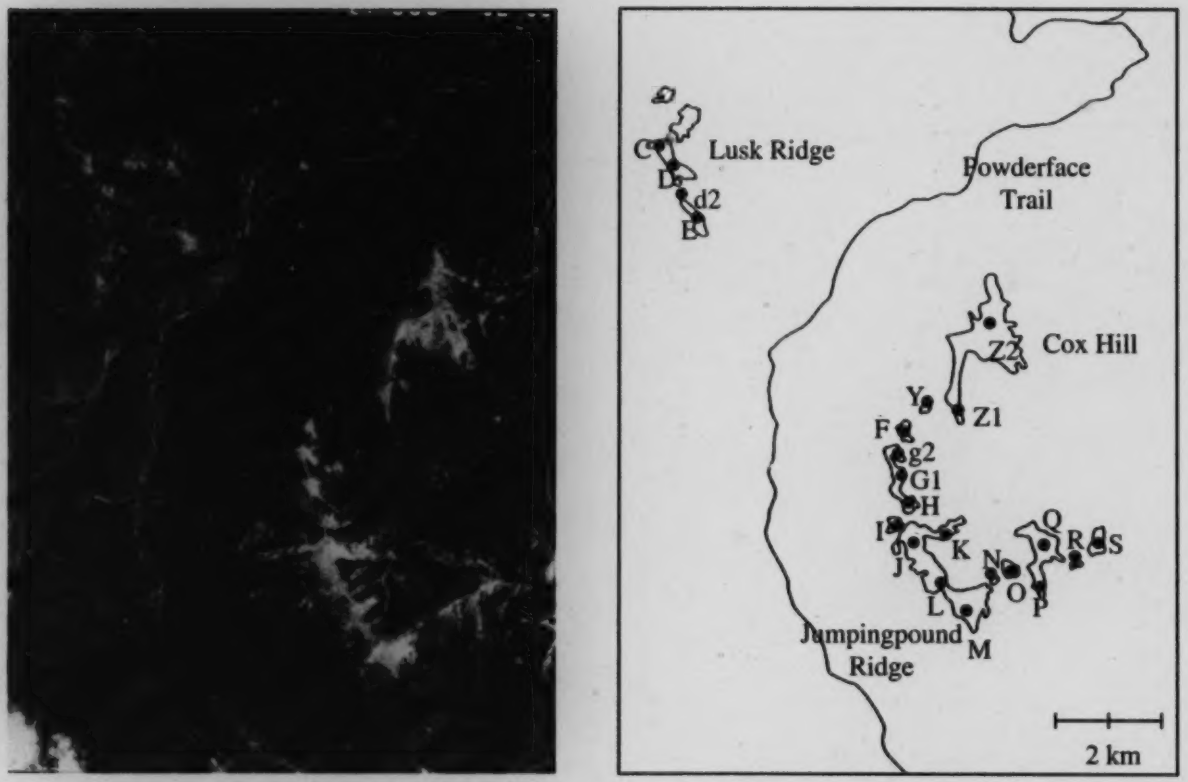


Figure 3-1. Study area. Left: 1993 aerial photograph (1: 40,000). Right: Schematic represenation of high-altitude (>2000m) meadows in the study area, showing locations of the twenty-two sample sites.

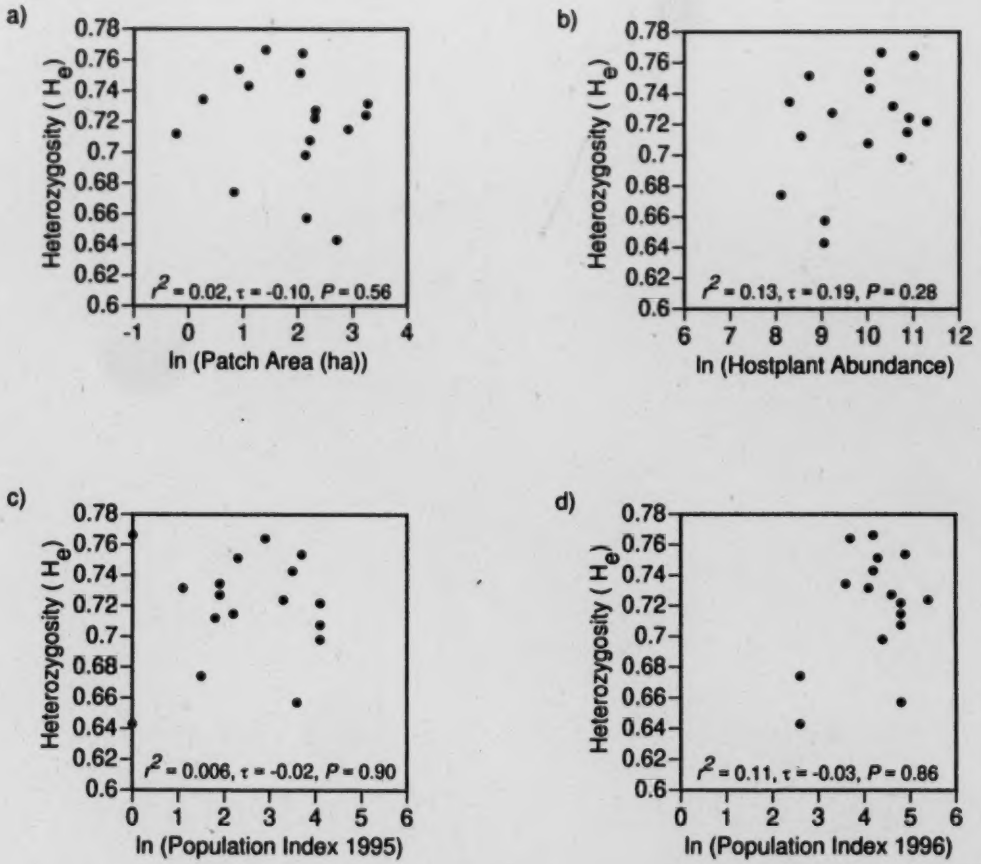


Figure 3-2. Genetic variation (expected heterozygosity averaged over loci; H_e) versus various measures of population size. All sample sites are considered separately.

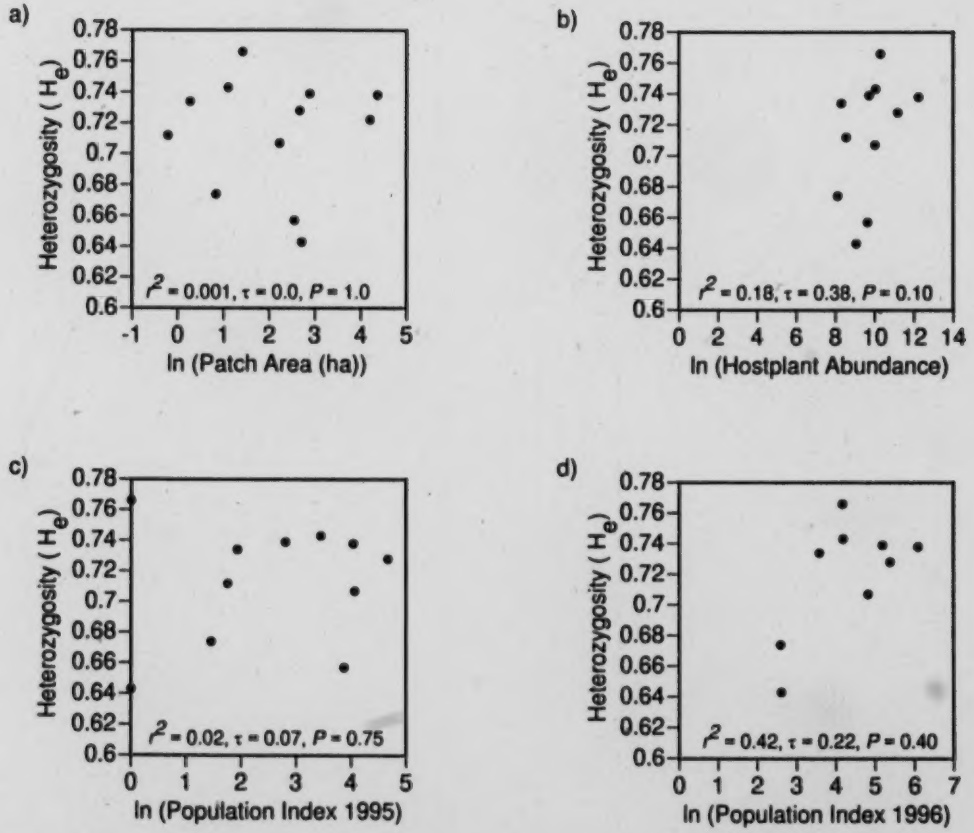


Figure 3-3. Genetic variation (expected heterozygosity averaged over loci; H_e) versus various measures of population size. Sample sites within a single, uninterrupted meadow are considered as one population - allele frequencies are averaged and measures of population size are summed over sites.

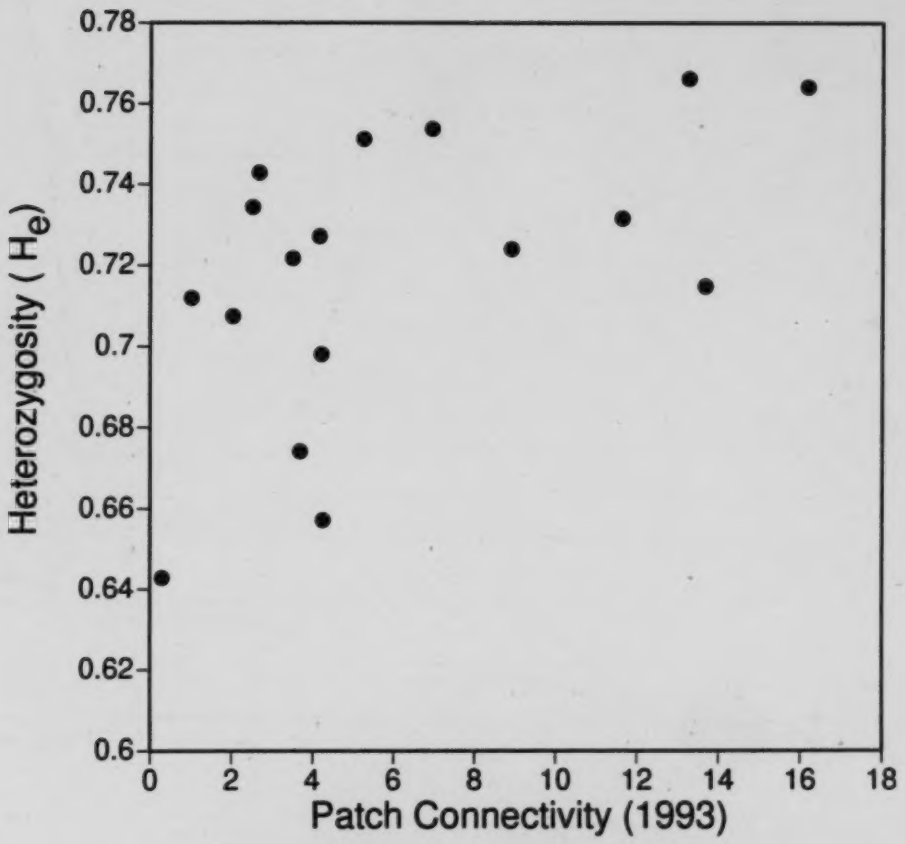


Figure 3-4. Genetic variation (expected heterozygosity averaged over loci; H_e) versus patch connectivity. The measure of connectivity takes into account differential effects of intervening forest versus meadow habitat. Distances and patch areas used to calculate connectivities were obtained from 1993 aerial photographs.

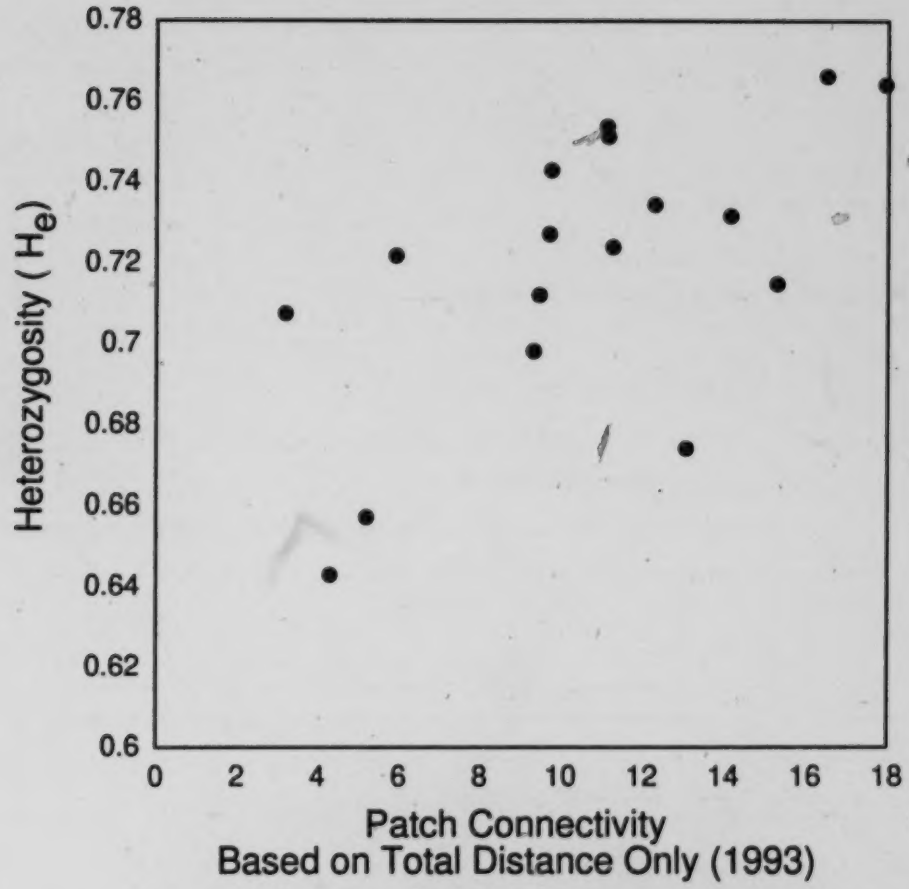


Figure 3-5. Genetic variation (expected heterozygosity averaged over loci; H_e) versus patch connectivity. The measure of connectivity does not take into account differential effects of intervening forest versus meadow habitat. Distances and patch areas used to calculate connectivities were obtained from 1993 aerial photographs.

Figure 3-6. Aerial photographs of Jumpingpound Ridge in 1952 and 1993 showing reforestation of burned areas and increasing forest cover over 40 years.

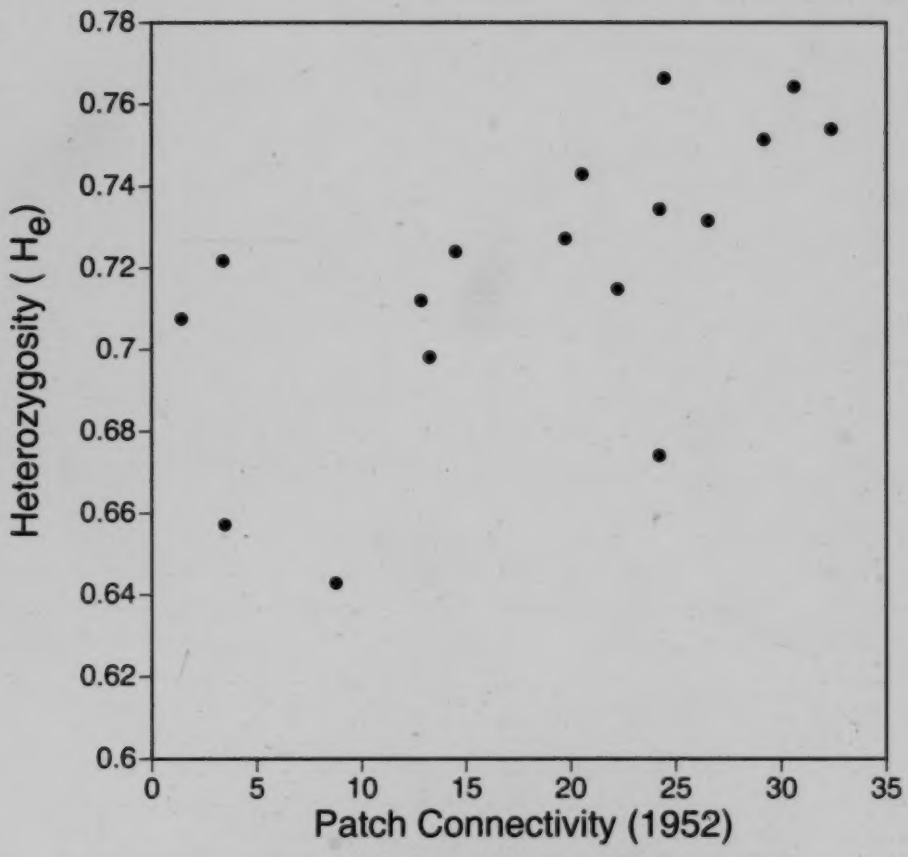


Figure 3-7. Genetic variation (expected heterozygosity averaged over loci; H_e) versus patch connectivity. The measure of connectivity takes into account differential effects of intervening forest versus meadow habitat. Distances and patch areas used to calculate connectivities were obtained from 1952 aerial photographs.

Chapter 4:

Large-scale patterns of population differentiation and landscape connectivity in the alpine butterfly, *Parnassius smintheus*

Introduction

Gene flow is a fundamental microevolutionary force that can alter allele frequencies within populations; however, its role in evolution has been highly controversial. Wright (1977) and Fisher (1930) disagreed on whether restricted gene flow promotes adaptation within species, sparking a debate that continues today (Coyne *et al.* 1997; Wade & Goodnight 1998). While Wright (1977) considered restricted gene flow and genetic drift in local populations to be beneficial for adaptation, Fisher (1930) argued that adaptation most likely occurs in large, panmictic populations. The role of restricted gene flow in population differentiation and speciation has been no less contentious. Mayr (1942) viewed gene flow as the primary factor maintaining cohesiveness within species and considered its interruption to be an important step in speciation. In contrast, Ehrlich & Raven (1969) were highly skeptical of its influence in maintaining species cohesiveness, attributing a greater importance to selection. In the past four decades, there has been a proliferation of studies describing the genetic structure of natural populations. From these, it has become apparent that the degree of genetic differentiation among local populations, and the levels and spatial extent of gene flow, vary widely among species and even within species (e.g., Britten *et al.* 1995). Thus, Slatkin (1985, 1987) has suggested that there is probably no single role of gene flow in adaptation and differentiation, but that its role varies depending on its strength relative to other evolutionary forces and on the geographic distribution of populations. The relative strength of gene flow may be affected by a variety of ecological factors that vary both within and among species such as dispersal ability, dietary

specialization, phenological asynchrony among populations, habitat patchiness, habitat persistence, and population persistence (Peterson & Denno 1998a). Elucidating the factors that affect the levels and spatial extent of gene flow is clearly a fundamental first step in understanding its role in evolution.

Though considerable effort has been invested in describing genetic structure of natural populations, the influence of ecological factors on gene flow and population differentiation has received far less attention and is relatively poorly understood (Peterson & Denno 1997). Of the ecological factors mentioned above, the effects of dispersal ability on genetic structure have been most thoroughly reviewed. Peterson & Denno (1998a) and Bohonak (1999) have found, across a wide range of species, that dispersal ability is generally positively correlated with inferred levels of gene flow. Variation in dispersal ability has also been shown to affect population differentiation within a species (Peterson & Denno 1997). However, Bohonak (1999) points out a number of studies in which dispersal ability and levels of gene flow do not correspond. In addition, it is not uncommon for species with apparently low dispersal ability to display high gene flow, though the reverse situation is much less frequently observed (Slatkin 1987).

A number of studies have also examined the effects of habitat connectivity on levels of gene flow and population differentiation, both within species (King 1987; Young *et al.* 1993; Britten *et al.* 1995; Johnson & Black 1995; VanDongen *et al.* 1998; Hutchison & Templeton 1999; Knutsen *et al.* 2000) and among species (Caccone & Sbordoni 1987; Shoemaker & Jaenike 1997; Peterson & Denno 1998b). Recently, many such studies have been motivated by concern over the potential impacts of habitat fragmentation on the genetic variation and evolutionary potential of species (Young *et al.* 1993; VanDongen *et al.* 1998; Knutsen *et al.* 2000). In many cases, greater habitat continuity in the landscape is associated with higher inferred gene flow. However, exceptions to this general pattern exist. For example, Young *et al.* (1993) found apparently higher levels of gene flow in fragmented populations of sugar maple, *Acer saccharum*, than in continuous populations;

they suggested that this may have been due to higher wind speed and therefore, greater wind-assisted dispersal of pollen in the open, fragmented landscape. Clearly, ecological factors may affect gene flow in complex and sometimes unanticipated ways.

Here, we examine how variation in landscape connectivity affects population differentiation and inferred levels of gene flow in the butterfly *Parnassius smintheus* Doubleday (1847). This species occurs in mountainous environments in western North America, inhabiting alpine and subalpine meadows. In our study area in the southern Canadian Rocky Mountains, populations are most commonly found just at or above tree-line. Movement of *P. smintheus* is generally restricted. Using mark-recapture methods Roland *et al.* (2000) found that the average dispersal distance of adults was approximately 150 m and the maximum dispersal distance observed was 1729 m. Movement is also strongly influenced by intervening landscape, and forested areas are much more formidable barriers to movement than are open meadows (Roland *et al.* 2000). In this species, restricted dispersal appears to be translated into restricted gene flow. Using highly variable microsatellite DNA markers, we have found significant genetic structure, in the form of isolation by distance, at the same spatial scale as the mark-recapture study with no more than 12 km separating the most distant populations (Keyghobadi *et al.* 1999). We have also found that, as for dispersal, gene flow among local populations is impeded to a greater extent by intervening forests than by intervening, open meadows (Keyghobadi *et al.* 1999).

In this study, we determine whether the effects of intervening landscape on dispersal and genetic differentiation that we have observed at a very fine spatial scale are also apparent at a larger spatial scale (up to 58 km separating sites). In the small scale study of population genetic structure (Keyghobadi *et al.* 1999), the study sites were distributed as a chain along ridge-tops in a one-dimensional configuration. Because movement of *P. smintheus* is largely restricted to ridge-tops, we could easily measure distances between sites along the ridges and divide those distances into the different landscape elements of forest or open meadow. The relative influences of these landscape

components on genetic distances between sites were then examined using Mantel and partial Mantel tests (Mantel 1967). This analytical approach is not applicable to a larger spatial scale where the sampled sites are distributed in a two-dimensional configuration. With sites distributed in two dimensions, there is no single, most probable path of movement between any two sites and division of distances between sites into different landscape elements is no longer straightforward.

To analyze the effects of landscape on genetic differentiation and gene flow at a larger spatial scale, we compare genetic structure in three regions differing in the connectivity of open habitats above tree-line. In particular, we focus on comparing patterns of isolation by distance among the three regions. Comparison of patterns of isolation by distance is a very useful approach to assessing the effects of ecological factors, such as landscape connectivity or dispersal ability, on population genetic structure (Peterson & Denno 1998a). This approach allows one to compare the relative influences of gene flow and genetic drift over various geographical distances (Hutchison & Templeton 1999) and is more informative than a single regional estimate of population structure or gene flow. In our study, comparison of genetic structure among the regions is facilitated by sampling similar numbers of populations over very similar spatial scales in each region.

One particular strength of this study is the ability to compare levels of differentiation and patterns of isolation by distance observed at the large scale to those observed at a much finer scale (Keyghobadi *et al.* 1999). This is a significant aid to interpretation of results at the large scale because the spatial scale of sampling can have a considerable impact on observed patterns of genetic variation (Peterson & Denno 1998a; Pogson *et al.* 2001).

Materials and Methods

Study area

Our study area encompassed 27 sampling sites in the foothills and front ranges of the Canadian Rocky Mountains, in Banff National Park and Kananaskis Country, Alberta (Fig. 4-1, Table 4-1). All sites were meadows, at or above tree-line, containing the larval host plant, stonecrop (*Sedum lanceolatum*).

The sampling sites can be divided into three regions, representing two very different landscape types. Eight sites were located in the region we refer to as East Kananaskis, which represents the landscape of the foothills. This is a hilly landscape dominated by forests, where open habitats above tree-line are relatively small and isolated. Three sample sites in East Kananaskis (E, Q, and Z) were also part of the fine-scale study of population genetic structure (Keyghobadi *et al.* 1999; Chapter 3). Nine and ten sample sites, respectively, were located in West Kananaskis and Banff, which represent the landscape of the mountain front ranges. This is a mountainous landscape that displays greater topographical relief than the foothills and is less heavily forested. Here, open habitats above tree-line are much larger and better connected.

Measurement of landscape variables

Geographical coordinates (degrees, minutes) of sampling sites were determined from 1:250,000 topographical maps and distances between sites were calculated from those coordinates using the GeoDist program of the R-PACKAGE (Legendre & Vaudor 1991). Within each region, the relative coverage of high-altitude (>2000 m), open (*i.e.* non-forested) habitat was quantified from 1:250,000 topographical maps that showed forested areas (Dept. of Energy, Mines, and Resources. Canada. 1991). A rectangular quadrat encompassing all of the sampling sites was delimited for each region, with the four edges placed such that they extended 2.5 km past the furthest outlying site in each direction (Fig. 4-1). Because our sampling sites lay predominantly in a Northwest-Southeast direction,

reflecting the directions of the main ridges and valleys in the area, the rectangular quadrats also lay in this direction. The areas of all patches of high-altitude, non-forested habitat within each such quadrat were traced onto Mylar paper, digitally scanned, and quantified using the image-analysis software IMAGETOOL (Wilcox *et al.* 1995). The proportion of each quadrat covered by such habitat was used as an index of the connectivity of the region for *P. smintheus*. The mean area of individual patches of high-altitude, open habitat within each quadrat was also used as a measure of landscape connectivity. In some cases, patches at the edges of a quadrat were actually parts of larger patches that mostly lay outside of the quadrat. These presented a number of problems for calculating mean patch area. If the area of the entire large patch was used, then the boundaries of the analyzed region would be extended far beyond the sampled region. However, if only the portion of the patch lying within the quadrat was used, then we would be underestimating patch areas. In addition, removing all patches that were not entirely within a quadrat from the analysis would have meant removing some large patches that lay mostly but not completely within the quadrats, and that clearly contributed to the connectivity of the sampled region. Therefore, to balance these considerations, in calculating the mean patch area for a quadrat, only those patches for which > 50% of the area lay within the quadrat were counted and only the portion lying within the quadrat was used. Not all high-altitude open areas are necessarily habitat for *P. smintheus*. Such areas also include meadows in which the butterfly or its host plant are not found (Fownes 1999), as well as bare rock, talus, and scree. Therefore, the landscape variables we measured are not indices of habitat connectivity per se, but they are indices of the connectivity of the landscape with respect to movement of *P. smintheus*.

Sample collection and microsatellite genotyping

Tissue samples from adult butterflies were collected in 1995, 1996, and 1999 from the 27 sites (Table 4-1). Samples were either small wing clippings (approx. 0.15 cm^2) or whole adult butterflies. All samples were placed individually in glassine envelopes and stored at -

80 °C upon return from the field. Genomic DNA was isolated from samples with the QIAamp™ tissue extraction kit (QIAGEN). For wing clippings, the entire sample was used. For whole butterflies, approximately 25 mg of abdominal tissue or thoracic tissue was used for males and females, respectively. Each sample was typed at seven microsatellite loci (Ps50, Ps81, Ps85, Ps76, Ps163, Ps165, and Ps262) as described previously (Keyghobadi *et al.* 1999; Chapter 6), using one fluorescently labeled primer per locus (labeled with 6-FAM, HEX, or TET; Applied Biosystems). Products of polymerase chain reaction (PCR) amplification were electrophoresed and detected on an Applied Biosystems 373A Automated Sequencer and analyzed using GENESCAN and GENOTYPER software (Applied Biosystems).

Data analysis

For each locus at each site, genotype frequencies were tested for conformity to Hardy-Weinberg expectations with the testing procedure of the program GENEPOP, version 3.1d (Raymond & Rousset 1995). Because we found evidence for null alleles at all loci (Keyghobadi *et al.*, 1999; Chapter 6), for each locus at each site a maximum likelihood estimate of the frequency of the null allele was calculated, and the frequencies of all other alleles were simultaneously re-estimated, using the estimation-maximization (EM) algorithm (Ceppellini *et al.* 1955; Yasuda & Kimura 1968; Long *et al.* 1995) as in Keyghobadi *et al.* (1999) [calculator available at <http://www.biology.ualberta.ca/jbrzusto/nullele>]. All further calculations of genetic structure and genetic variation were performed using these estimates of allele frequencies, including that of the null allele.

For each locus separately, we estimated F_{st} for each region using the Analysis of Molecular Variance (AMOVA) procedure of the program ARLEQUIN (Schneider *et al.* 2000). The program calculates F_{st} as described by Weir & Cockerham (1984). Because these calculations were based on allele frequencies and not genotype frequencies, the

AMOVA does not include a within-individual level of variance and the analysis assumes random union of gametes within populations. We feel that this is a valid assumption, as various aspects of the behaviour and life-history of this species suggest that mating within local populations is probably random (Chapter 6). For each region, F_{st} estimates from different loci were combined by weighted averaging as recommended by Weir & Cockerham (1984). The program ARLEQUIN tests the significance of F_{st} for each locus by randomly permuting alleles among populations. We combined the probability values from each of these significance tests using the meta-analytic approach described by Sokal & Rohlf (1995; pp.794-797), to obtain the probability of significance of the overall F_{st} (combined over loci) for each region.

Nei's standard genetic distance (Nei 1972) was calculated between all pairs of sites within regions [calculator at <http://www.biology.ualberta.ca/jbrzusto>]. To test for isolation by distance in each region, the significance of the correlation between genetic distance and geographic distance was determined using the Mantel test (Mantel 1967), as executed by the R-PACKAGE program (Legendre & Vaudor 1991). For each test, 3000 matrix randomizations were used. We chose to use Nei's standard genetic distance (D_S) because, in comparing various genetic distance measures for use with microsatellite data, Paetkau *et al.* (1997) found that D_S performed very well for detecting isolation by distance at a fine geographic and temporal scale, displaying relatively low variance.

Unbiased estimates of expected heterozygosity for each locus within each site were calculated as $(1-\Sigma p_i^2)(2N/2N-1)$, where p_i is the estimated frequency of allele i , and N is the number of individuals sampled (Nei & Roychoudhury 1974). For each site, expected heterozygosities were averaged across loci. Differences in within-site average expected heterozygosity among the regions were tested using pairwise Mann-Whitney U tests (Zar 1984; pp. 138-146) with a Bonferroni correction for total experimentwise error.

Comparison to genetic structure at a smaller spatial scale

We have previously studied population genetic structure of *P. smintheus* at a much finer spatial scale (Keyghobadi *et al.* 1999; Chapter 3), and in conjunction with a mark-recapture study (Roland *et al.* 2000). Seventeen sites separated by distances of 0.15 to 12 km were sampled and individuals were typed at the same seven microsatellite loci described above (Ps50, Ps81, Ps85, Ps76, Ps163, Ps165, and Ps262; Keyghobadi *et al.* 1999; Chapter 3; Chapter 6). These sites were located along three ridges in the East Kananaskis region (Fig. 4-1).

As for sites in the larger scale study area, estimates of allele frequencies, including a null allele, at all loci were obtained using the EM algorithm. These allele frequencies were used to calculate estimates of population genetic structure for the fine-scale study area, as was done for each of the regions in the larger scale study. F_{st} among the 17 sampling sites was estimated as described above and isolation by distance was examined by testing for a correlation between Nei's standard genetic distance (Nei 1972) and geographic distance using the Mantel test (Mantel 1967).

Results

Landscape connectivity

The proportion of area covered by high-altitude, open habitat was 18.7% in East Kananaskis, 53.9% in Banff, and 52.9% in West Kananaskis (Table 4-2). The average size of individual patches of high-altitude, open habitat was 2.74 km^2 in East Kananaskis, 47.11 km^2 in Banff, and 57.44 km^2 in West Kananaskis (Table 4-2). Thus, the connectivity of the landscape with respect to movement of *P. smintheus* was considerably lower in the foothills landscape (East Kananaskis) than in the front ranges (Banff and West Kananaskis). In addition, based on the average size of individual patches of high-altitude non-forested habitat, the West Kananaskis region appeared to have slightly higher connectivity for *P. smintheus* than did the Banff region.

Hardy-Weinberg equilibrium

For the 27 sites in the larger scale study area, a total of 189 tests of conformity to Hardy-Weinberg proportions were performed. Of these, 131 showed significant deviations from expected proportions. Even when the total experimentwise error rate was controlled by a conservative Bonferroni adjustment (α=.05/189=0.0003), 92 tests were significant. Deviations were observed at all seven loci and always involved excess homozygosity. Homozygote excess at these loci has been observed previously in samples from the fine scale study area and has been attributed to the occurrence of null alleles (Keyghobadi *et al.* 1999; Chapter 6). Null alleles are therefore the most probable cause of homozygote excess in the large scale study sites as well. Homozygote excess due to assortative mating of related individuals has previously been ruled out based on various aspects of the life history and mating behaviour of this species (Chapter 6). In the small scale study, excess homozygosity due to a Wahlund effect, or the erroneous combination of separate populations in a single sample, was also ruled out based on the very fine spatial scale of sampling (Chapter 6). At the large scale, the possible occurrence of a Wahlund effect cannot be dismissed as easily. However, excess homozygosity within sites, measured as the difference between observed and expected homozygosity, averaged over loci, did not differ between the large scale study sites (excluding sites E, Q, and Z) and the small scale study sites (Mann-Whitney U test: Z = -0.66, P = 0.51). This suggests that no additional factor other than null alleles is necessary to explain the excess homozygosity observed in the large scale study sites.

Microsatellite variability

Overall, we detected high microsatellite variability in these populations. The average number of alleles within samples ranged from 6.7 to 13.0, and the average expected heterozygosity ranged from 0.65 to 0.79 (Table 4-1).

Genetic structure and F_{st} estimates

Estimates of F_{st} varied among the three regions and were correlated with the connectivity of high altitude, open habitat (Table 4-2). Thus, F_{st} was 0.038 ($P < 0.001$) in East Kananaskis which was the most fragmented region, 0.011 ($P < 0.001$) in Banff, and 0.0016 ($P = 0.13$) in West Kananaskis which was the least fragmented region. Estimated F_{st} for West Kananaskis was comparable to the value observed in the fine scale study area ($F_{st} = .0017$; $P = 0.18$).

Isolation by distance

We observed a significant, positive correlation between D_s and geographic distance in West Kananaskis ($r = 0.59$, $P = 0.0007$; Fig. 4-2) and Banff ($r = 0.50$, $P = 0.03$; Fig. 4-3), but not in East Kananaskis ($r = -0.28$, $P = 0.13$; Fig. 4-4). Within the fine-scale study area, which is located in the East Kananaskis region, we observed a significant positive correlation between D_s and geographic distance ($r = 0.54$, $P = 0.0003$; Fig. 4-5; Chapter 3). The range of D_s values observed in West Kananaskis was very similar to the range observed in the fine scale study area (Fig. 4-2, Fig. 4-5).

Differences in within-population genetic variation among regions

We found significant differences in average expected heterozygosity (H_e) among the three regions (Table 4-2). Pairwise Mann-Whitney U tests, with sequential Bonferroni correction of the α values (Sokal & Rohlf 1995; p. 241), showed that H_e was significantly lower in East Kananaskis than in both Banff ($Z = -3.11$, $P = 0.002$) and West Kananaskis ($Z = -2.41$, $P = 0.016$). H_e was slightly higher in Banff than in West Kananaskis (Table 4-2) and this difference was marginally significant ($Z = -1.96$, $P = 0.05$).

Discussion

We found that differences in landscape connectivity among three geographic regions were associated with differences in the degree of population differentiation and inferred levels of gene flow in the alpine butterfly, *P. smintheus* (Table 4-2). First, F_{st}, which indicates the degree of population subdivision, was lowest in the least fragmented region (West Kananaskis) and highest in the most fragmented region (East Kananaskis). Second, the strength of isolation by distance was highest in the least fragmented region and lowest in the most fragmented region, where no significant correlation was observed. We interpret these differences in patterns of isolation by distance as indicating increasing levels of gene flow with increasing landscape connectivity (discussed below). Also, within-population genetic variation was significantly lower in the most fragmented region than in the other two regions. Low within-population variation may be associated with population isolation, if genetic drift is not negligible (Chapter 3). Overall, our results indicate a positive relationship between inferred levels of gene flow and landscape connectivity. Our results also indicate that the effects of landscape on dispersal and genetic structure that have previously been observed in this species at a small spatial scale (up to 12 km: Keyghobadi *et al* 1999; Roland *et al.* 2000) are also apparent at a larger spatial scale (up to 58 km).

Gene flow versus genetic drift

For neutral genetic markers, differences in F_{st} and patterns of isolation by distance among geographic regions potentially reflect variation in either rates of genetic exchange among populations (gene flow), or levels of genetic drift within populations, or both. In general, the effects of gene flow and genetic drift on observed patterns of genetic variation are difficult to disentangle (Hutchison & Templeton 1999). Thus, although for convenience we refer to variation in levels of gene flow among the three regions, what we have really measured is variation in levels of gene flow relative to drift. However, because gene flow

is mediated by dispersal, information on individual movements can be used to make inferences about the role of gene flow, independent of drift, in determining observed patterns of genetic structure. In this case, the effects of landscape connectivity on genetic structure are in accord with observed effects of landscape on movement of individuals (Roland *et al.* 2000). That is, the degree of genetic differentiation among populations decreases with increasing connectivity of high-altitude, non-forested habitats. At the same time, movement of individuals occurs approximately twice as readily through open meadows than through forested areas (Roland *et al.* 2000). The concordance of these results strongly suggests that decreasing genetic differentiation among populations with increasing landscape connectivity must be due, at least partly, to increasing levels of gene flow in the more connected landscapes.

Lack of isolation by distance in the foothills landscape

A significant correlation between genetic similarity/dissimilarity and geographic distance, referred to as isolation by distance, is predicted by both continuous and stepping-stone models of population structure under conditions of restricted gene flow (Wright 1943; Kimura & Weiss 1964). With increasing geographic distance, the influence of gene flow relative to drift declines, such that both the mean and variance of genetic distances among populations increase. Lack of a significant pattern of isolation by distance over a particular geographical scale for neutral markers, as observed in East Kananaskis (Fig. 4-4), may occur for one of three reasons. First, populations may not be at equilibrium with respect to the forces of gene flow and drift following some historical perturbation. Second, gene flow may be relatively high over the entire distance range. In this case, one expects to observe both a low mean and a low variance in genetic distances (Hutchison & Templeton 1999). Third, gene flow may be relatively low over the distances sampled such that populations are essentially isolated and allele frequencies are affected mostly by drift. In

this case, one expects to observe a high mean and high variance in genetic distances (Hutchison & Templeton 1999).

We have no reason to suspect that the East Kananaskis region as a whole has experienced appreciably different large-scale historical perturbations than the other two regions to which it is geographically quite proximate (Fig. 4-1). Given that a significant pattern of isolation by distance was observed in both Banff and West Kananaskis, it seems unlikely that the lack of such a pattern in East Kananaskis would be due to non-equilibrium conditions.

Because of the sedentary nature of this species (Roland *et al.* 2000), lack of isolation by distance in East Kananaskis would appear to be due to low levels of gene flow over the distances sampled. However, it is not uncommon for apparently sedentary species to display surprisingly high levels of gene flow, perhaps due to occasional episodes of long-distance dispersal, extinction-colonization dynamics, or various historical factors (Slatkin 1985). Also, means and variances of genetic distance (D_S) were not considerably higher in East Kananaskis than in the other two regions (Figs. 2-4). However, F_{ST} was highest in East Kananaskis indicating a higher degree of population differentiation than in the other regions. Furthermore, the significant pattern of isolation by distance observed in the small scale study area (Fig. 4-5), which was located within East Kananaskis (Fig. 4-1), very strongly suggests that lack of isolation by distance at the larger scale was most likely due to low levels of gene flow. If lack of a significant correlation between genetic distance and geographic distance is due to high gene flow, then decreasing the spatial scale of analysis is not expected to reveal such a correlation. However, if lack of isolation by distance is due to severely restricted gene flow, then at some smaller spatial scale a pattern of isolation by distance should eventually become evident. Thus, our results suggest that populations in the foothills landscape are quite isolated from each other at the larger spatial scale, but are connected by gene flow at a smaller scale. Peterson & Denno (1998a) have suggested that in the many sedentary phytophagous insects for which isolation by distance

is not observed, analysis at a smaller spatial scale should reveal significant isolation by distance. Our results from East Kananaskis at both large and small scales substantiate this prediction.

Comparison of Banff and West Kananaskis regions

Although West Kananaskis and Banff both represent the front ranges landscape, comparison of genetic structure between these two regions also supported the hypothesis that higher gene flow is associated with higher landscape connectivity. The proportion of high-altitude open habitat in the two regions was very similar, but the average size of individual patches of such habitat was higher in West Kananaskis than in Banff (Table 4-2), mostly due to the presence of several small patches in Banff, such as the one from which the North Cascade 1 population was sampled (Fig. 4-1). Thus, landscape connectivity appeared to be slightly higher in West Kananaskis. In association with higher landscape connectivity, West Kananaskis displayed a lower F_{st} (Table 4-2), suggesting higher levels of gene flow. Comparison of patterns of isolation by distance also suggest higher gene flow in West Kananaskis. In West Kananaskis we observed a stronger correlation between genetic distance and geographic distance; for a given geographic distance, the variance in D_S was lower in West Kananaskis than in Banff (Fig. 4-2, Fig. 4-3). Thus, over most distances the effects of gene flow relative to drift appeared to be stronger in West Kananaskis. In addition, the slope of the regression of genetic distance versus geographic distance was higher in Banff than in West Kananaskis (Fig. 4-2, Fig. 4-3) such that in the Banff region there was a greater increase of genetic distance for a given increase in geographic distance. Therefore, a given unit distance in Banff was not equivalent to a given unit distance in West Kananaskis, presumably due to differences in landscape connectivity between the regions.

The differences in genetic structure and apparent levels of gene flow between West Kananaskis and Banff are quite noticeable and appear to be out of proportion to the

relatively small difference in landscape connectivity. One very likely explanation for this is that there are other factors affecting movement and gene flow of *P. smintheus* which we have not considered in our relatively simple landscape variables. For example, the Banff region contains a large lake and a town site (Fig. 4-1) that may impede movement of *P. smintheus*. No such barriers separate the sampled populations in West Kananaskis. Another important factor that we have not considered is the spatial configuration of sample sites, which differs between the regions. While sample sites in West Kananaskis lie largely along a single valley in a linear configuration, the configuration of sample sites in Banff is somewhat less linear and more sample sites are situated in an East-West direction from each other, perpendicular to the direction of the main ridges and valleys (Fig. 4-1). As a result, sites in West Kananaskis are separated by at most a single, large forested valley. However, despite similar geographic distances among sites in the two regions, some sites in Banff are separated by two large forested valleys (e.g., Cascade 1 and Forty Mile Creek). Thus, although the two regions may not differ greatly in landscape connectivity, because of differences in the spatial configuration of sample sites, the populations in Banff may be separated to a greater extent by barriers to movement (*i.e.* forested valleys) than populations in West Kananaskis.

Another possible explanation for the unexpectedly large differences in genetic structure between West Kananaskis and Banff is that there has been selection for reduced dispersal ability in the slightly more fragmented region of Banff, such that the effects of a slight difference in landscape structure have been amplified. Habitat fragmentation may exert selection pressures on dispersal ability or tendency (Van Dyck & Matthysen 1999). In theory, habitat fragmentation could lead to selection for either increased or decreased dispersal depending on the relative costs and benefits of dispersal (Van Dyck & Matthysen 1999). In flying insects, both trends have been noted (Taylor & Merriam 1995, Hill *et al.* 1999). If, in *P. smintheus*, there has been selection for increased dispersal in less connected landscapes, then the result has not been sufficient to completely compensate for

the negative correlation between landscape connectivity and levels of genetic differentiation among populations. However, it is interesting to contemplate the possibility that there has been selection for more sedentary behaviour in more fragmented landscapes. In that case, landscape connectivity and dispersal ability will be correlated, setting up a positive feedback that could amplify the effects of relatively small differences in landscape connectivity between regions. Thus, appreciable differences in genetic structure may be seen between regions with apparently small differences in landscape connectivity. Comparative studies of dispersal behaviour and flight morphology in adult *P. smintheus* in the different regions would shed light on the likelihood of this hypothesis.

Historical changes in landscape structure

Differences in population genetic structure among the three regions may reflect historical rather than contemporary differences in landscape structure or the geographic distribution of populations. On the time-scale of thousands of years, the distribution of populations may have been different during different climatic periods in the past, such as the Wisconsin glaciation (during which an ice-free corridor extended along the Canadian Rocky Mountains (Flint 1971)) or the Little Ice Age. On the time-scale of tens to hundreds of years, forest fires may have acted to increase the connectivity of non-forested habitats in the landscape.

It is unlikely that microsatellite variation would reflect patterns produced by climatic changes occurring thousands of years in the past. Because of their high mutation rate, variation at microsatellite loci is expected to respond relatively quickly to environmental change and to reflect more recent population structure (Sunnucks 2000). For example, in our small-scale study area, genetic differentiation among sites was correlated with the amount of forested habitat separating sites in 1993, even though the study area was largely unforested less than 60 years earlier following a large fire in 1938 (Keyghobadi *et al.* 1999, Chapter 3). This indicates a rapid response to changes in landscape structure. It is far

more likely that differences in recent fire history among the three regions may have influenced current patterns of population genetic structure as reflected by microsatellite variation. However, a relationship between fire history and population genetic structure is not apparent. In general, fire cycles are longer in the western regions of the Alberta Rockies (Rogeau 1996). Thus, we would expect fires to burn and increase the connectivity of non-forested habitats more often in the foothills landscape than the front ranges landscape. However, populations in the foothills landscape displayed the greatest degree of genetic differentiation. Thus, if fires have acted to increase landscape connectivity historically, their impact has not been sufficient to prevent isolation of populations in the foothills landscape. More detailed studies are necessary to assess the impacts of individual fires at known times in the past on current patterns of genetic variation (Chapter 3). However, there is currently no evidence that observed differences in genetic structure among the three regions are better explained by historical rather than contemporary differences in landscape structure.

Low estimated values of F_{st} and high variability of microsatellites

Overall, we observed relatively low values of F_{st} within all three regions, in contradiction to the apparently sedentary nature of this species (Roland *et al.* 2000). The commonly used result of Wright's island model of migration, $F_{st} = 1/(1 + 4\ Nm)$, yields estimates of gene flow (the product 'Nm' where N is effective population size and m is the rate of migration) of 6, 22, and 156 in East Kananaskis, Banff, and West Kananaskis, respectively. These values seem extraordinarily high given the low rates and distances of dispersal observed in this species (Roland *et al.* 2000), and given that a significant pattern of isolation by distance has been detected over a small spatial scale (Keyghobadi *et al.* 1999). The low values of F_{st} we observed may be attributable to the high variability of the microsatellite loci we used. Slatkin (1995) has pointed out that because of the high variability and high mutation rates of microsatellites, estimates of F_{st} based on microsatellite data tend to underestimate

the degree of population differentiation. Paetkau *et al.* (1997) and Hedrick (1999) have also proposed that for highly variable markers, the magnitude of differentiation among sites is negatively correlated with levels of diversity within sites. High estimates of gene flow in species that are generally considered sedentary (e.g., Cullenward *et al.* 1979) have, as mentioned previously, been attributed to infrequent episodes of long-distance dispersal, extinction-colonization dynamics, or historical factors (Slatkin 1985). However, in the case of *P. smintheus*, there is no reason to invoke such factors as the low estimated values of F_{st} are most parsimoniously explained by the high variability and high mutation rates of the molecular markers used. Thus, although the values of F_{st} that we have measured are informative for comparing levels of population differentiation among the three regions, they cannot be used to obtain estimates of gene flow (Nm). Regardless of other problems associated with estimation of levels of gene flow (Nm) from F_{st} using Wright's island model (Whitlock & McCauley 1999), the high mutation rate and high variability of the microsatellite markers we used appear to lead to spuriously high values of Nm.

Slatkin (1995) has designed an alternative measure of population differentiation specifically for microsatellite loci (R_{st}). Unfortunately, this measure is not necessarily more appropriate for our data. R_{st} uses information on allele sizes and assumes a generalized step-wise mutation model. At several of the *P. smintheus* microsatellite loci used here, size variation occurs in the flanking regions (Chapter 6). For example, several dinucleotide loci have alleles that differ in size by a single base-pair due to indels in the flanking regions (Chapter 6). Therefore, estimates of genetic structure such as R_{st} that are based on allele sizes are probably not applicable for most of the microsatellite loci we used.

Conclusion

Our results indicate that in *P. smintheus* levels of gene flow relative to drift vary with landscape structure and spatial scale. The relative strength of gene flow increased with increasing connectivity of the high-altitude, non-forested habitat. In the most fragmented

region, the relative strength of gene flow also increased with a five-fold decrease in spatial scale. Therefore, the role played by gene flow in adaptation and differentiation may vary among groups of populations in this species depending on the nature of the surrounding landscape and on the geographic distribution of populations.

Literature Cited

Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. *Quarterly Review of Biology* **74**: 21-45.

Britten, H. B., P. F. Brussard, D. D. Murphy and P. R. Ehrlich. 1995. A test for isolation-by-distance in Central Rocky Mountain and Great Basin populations of Edith's Checkerspot butterfly (*Euphydryas editha*). *Journal of Heredity* **86**: 204-210.

Caccone, A. and V. Sbordoni. 1987. Molecular evolutionary divergence among North American cave crickets. I. Allozyme variation. *Evolution* **41**: 1198-1214.

Ceppellini, R., M. Siniscalco and C. A. B. Smith. 1955. The estimation of gene frequencies in a random-mating population. *Annals of Human Genetics* **20**: 97-115.

Coyne, J. A., N. H. Barton and M. Turelli. 1997. A critique of Sewall Wright's shifting balance theory of evolution. *Evolution* **51**: 643-671.

Cullenward, M. J., P. R. Ehrlich, R. R. White and C. E. Holdren. 1979. The ecology and population genetics of an alpine checkerspot butterfly, *Euphydryas anicia*. *Oecologia* **38**: 1-12.

Ehrlich, P. R. and P. H. Raven. 1969. Differentiation of populations. *Science* **165**: 1228-1232.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Flint, R. F. 1971. Glacial and Quaternary Geology. John Wiley & Sons, Inc., New York.

Fownes, S. L. 1999. Meadow suitability and quality for the alpine butterfly, *Parnassius smintheus*, in the east slopes of the Canadian Rockies. M.Sc. Thesis. University of Alberta.

Hedrick, P. W. 1999. Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313-318.

Hill, J. K., C. D. Thomas and O. T. Lewis. 1999. Flight morphology in fragmented populations of a rare British butterfly, *Hesperia comma*. *Biological Conservation* **87**: 277-283.

Hutchison, D. W. and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**: 1898-1914.

Johnson, M. S. and R. Black. 1995. Neighbourhood size and the importance of barriers to gene flow in an intertidal snail. *Heredity* **75**: 142-154.

Keyghobadi, N., J. Roland and C. Strobeck. 1999. Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* **8**: 1481-1495. **Chapter 2 of this volume.**

Kimura, M. and G. H. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**: 561-576.

King, P. S. 1987. Macro- and microgeographic structure of a spatially subdivided beetle species in nature. *Evolution* **41**: 401-416.

Knutsen, H., B. A. Rukke, P. E. Jorde and R. A. Ims. 2000. Genetic differentiation among populations of the beetle *Bolitophagus reticulatus* (Coleoptera: Tenebrionidae) in a fragmented and a continuous landscape. *Heredity* **84**: 667-676.

Legendre, P. and A. Vaudor. 1991. The R package: multidimensional analysis, spatial analysis. Version 3.0. Université de Montréal, Montréal.

Long, J. C., R. C. Williams and M. Urbanek. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. *American Journal of Human Genetics* **56**: 799-810.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209-220.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York.

Nei, M. 1972. Genetic distance between populations. *American Naturalist* **106**: 283-292.

Nei, M. and A. K. Roychoudhury 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* **76**: 379-390.

Paetkau, D., L. P. Waits, P. L. Clarkson, L. Craighead, and C. Strobeck. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* **147**: 1943-1957.

Peterson, M. A. and R. F. Denno. 1997. The influence of intraspecific variation in dispersal strategies on the genetic structure of planthopper populations. *Evolution* **51**: 1189-1206.

Peterson, M. A. and R. F. Denno. 1998a. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *American Naturalist* **152**: 428-446.

Peterson, M. A. and R. F. Denno. 1998b. Life-history strategies and the genetic structure of phytophagous insect populations, pp. 263-322. In S. Mopper and S. Y. Strauss (eds.), Genetic Structure and Local Adaptation in Natural Insect Populations. Chapman & Hall, New York.

Pogson, G. H., C. T. Taggart, K. A. Mesa and R. G. Boutilier. 2001. Isolation by distance in the atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution* **55**: 131-146.

Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.

Rogeau, M-P. 1996 Understanding age-class distributions in the Southern Canadian Rockies. M.Sc. Thesis. University of Alberta.

Roland, J., N. Keyghobadi and S. Fownes. 2000. Alpine *Parnassius* butterfly dispersal: effects of landscape and population size. *Ecology* **81**: 1642-1653.

Schneider, S., D. Roessli and L. Excoffier. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. University of Geneva, Geneva.

Shoemaker, D. D. and J. Jaenike. 1997. Habitat continuity and the genetic structure of *Drosophila* populations. *Evolution* **51**: 1326-1332.

Slatkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology & Systematics* **16**: 393-430.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* **236**: 787-792.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457-462.

Sokal, R. R. and F. J. Rohlf. 1995. Biometry. W. H. Freeman and Company, New York.

Taylor, P. D. and G. Merriam. 1995. Wing morphology of a forest damselfly is related to landscape structure. *Oikos* **73**: 43-48.

Van Dyck, H. and E. Matthysen. 1999. Habitat fragmentation and insect flight: a changing 'design' in a changing landscape? *Trends in Ecology & Evolution* **14**: 172-174.

VanDongen, S., T. Backeljau, E. Matthysen and A. A. Dhondt. 1998. Genetic population structure of the winter moth (*Operophtera brumata* L.) (Lepidoptera, Geometridae) in a fragmented landscape. *Heredity* **80**: 92-100.

Wade, M. J. and C. J. Goodnight. 1998. The theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. *Evolution* **52**: 1537-1553.

Weir, B. S. and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

Whitlock, M. C. and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. *Heredity* **82**: 117-125.

Wilcox, D., B. Dove, D. McDavid, and D. Greer. 1995. ImageTool for Windows. Version 2.00. The University of Texas Health Science Center, San Antonio.

Wright, S. 1943. Isolation by distance. *Genetics* **28**: 114-138.

Wright, S. 1977. Evolution and the Genetics of Populations. Vol. 3. University of Chicago Press, Chicago.

Yasuda, N. and M. Kimura. 1968. A gene-counting method of maximum likelihood for estimating gene frequencies in ABO and ABO-like systems. *Annals of Human Genetics* **31**: 409.

Young, A. G., H. G. Merriam and S. I. Warwick. 1993. The effects of forest fragmentation on genetic variation in *Acer saccharum* Marsh. (sugar maple) populations. *Heredity* **71**: 277-289.

Zar, J. H. 1984. Biostatistical Analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Table 4-1. Genetic variation within sample sites. Sample size (N) refers to the number of individuals sampled. Avg. No. Alleles is the number of observed alleles, including a null allele, averaged over loci. Avg. He is the unbiased estimate of expected heterzygosity, (Nei & Roychoudhury 1974) averaged over loci.

Sample site	No. on map (Fig. 1)	Region	Sample size (N)	Avg No. Alleles	Avg He
E (Lusk Ridge)	1	East Kan.	40	9.7	0.708
Q (Jumpingpound Ridge)	2	East Kan.	40	10.6	0.727
Z (Cox Hill)	3	East Kan.	41	10.1	0.722
Moose Mtn.	4	East Kan.	26	8.7	0.736
Mt. Baldy	5	East Kan.	24	6.7	0.647
Powderface Ridge	6	East Kan.	31	10.0	0.773
Forget-Me-Not Ridge	7	East Kan.	34	11.0	0.742
Volcano Ridge	8	East Kan.	40	11.6	0.736
Wedge	9	West Kan.	40	10.4	0.744
Mt. Kidd	10	West Kan.	24	9.7	0.766
Mt. Allan	11	West Kan.	17	8.6	0.765
Fortress Mtn.	12	West Kan.	40	10.1	0.788
Mt. Kent	13	West Kan.	40	10.4	0.772
Ptarmigan Cirque	14	West Kan.	21	10.0	0.760
Elk Range	15	West Kan.	11	8.3	0.727
Mist Mtn.	16	West Kan.	31	13.0	0.776
Pigeon Mtn.	17	West Kan.	19	8.4	0.747
Mt. Peechee	18	Banff	18	6.9	0.749
North Cascade 1	19	Banff	15	6.9	0.788
Cascade 3	20	Banff	20	6.7	0.794
Cascade 1	21	Banff	40	8.3	0.790
Cascade 4	22	Banff	39	8.4	0.769
Flint Peak	23	Banff	40	9.1	0.781
FortyMile Creek	24	Banff	26	8.0	0.773
Stony Creek	25	Banff	40	9.7	0.763
Snow Creek	26	Banff	40	10.1	0.788
Panther Mtn.	27	Banff	40	8.4	0.768

Table 4-2. Landscape characteristics and population genetic structure in each of the sampled regions. Non-forested habitat indicates the proportionate coverage of high-altitude (>2000 m), non-forested habitat in a quadrat encompassing all sample sites within the region (Fig. 1). Avg. patch size is the mean area of patches of high-altitude (>2000 m), non-forested habitat in the same quadrat. For isolation by distance, *r* is the correlation coefficient for Nei's standard genetic distance (Nei 1972) and geographic distance, and *P* is the significance of the correlation as determined by a Mantel test. Avg. H_e is the unbiased expected heterozygosity (Nei & Roychoudhury 1974) averaged over loci within sites and then averaged over sites within each region.

				Genetic structure		Isolation by distance		
Region	Landscape type	Non-forested habitat	Avg. patch size (km^2)	F_{st}	*(P)*	*r*	*(P)*	Avg. He (±SD)
East Kananaskis	Foothills	18.7%	2.74	0.038	<.001	-0.28	0.13	0.724 (±0.036)
Banff	Front ranges	53.9%	47.11	0.011	<.001	0.50	0.03	0.776 (±0.019)
West Kananaskis	Front ranges	52.9%	57.44	0.0016	0.13	0.59	0.0007	0.760 (±0.014)

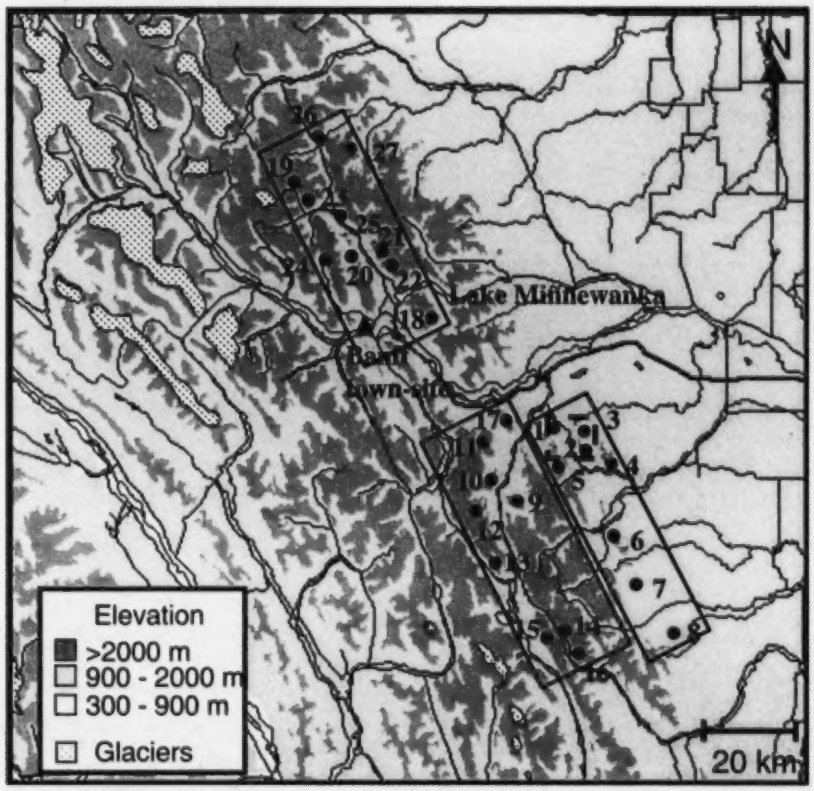


Figure 4-1. Locations of sample sites. Sites 1-8 are in East Kananaskis, sites 9-17 are in West Kananaskis, and sites 18-27 are in Banff. The three large rectangles indicate the quadrats used to measure landscape structure for each region. The small, dashed rectangle designates the location of the small-scale study area (Keyghobadi *et al.* 1999; Roland *et al.* 2000).

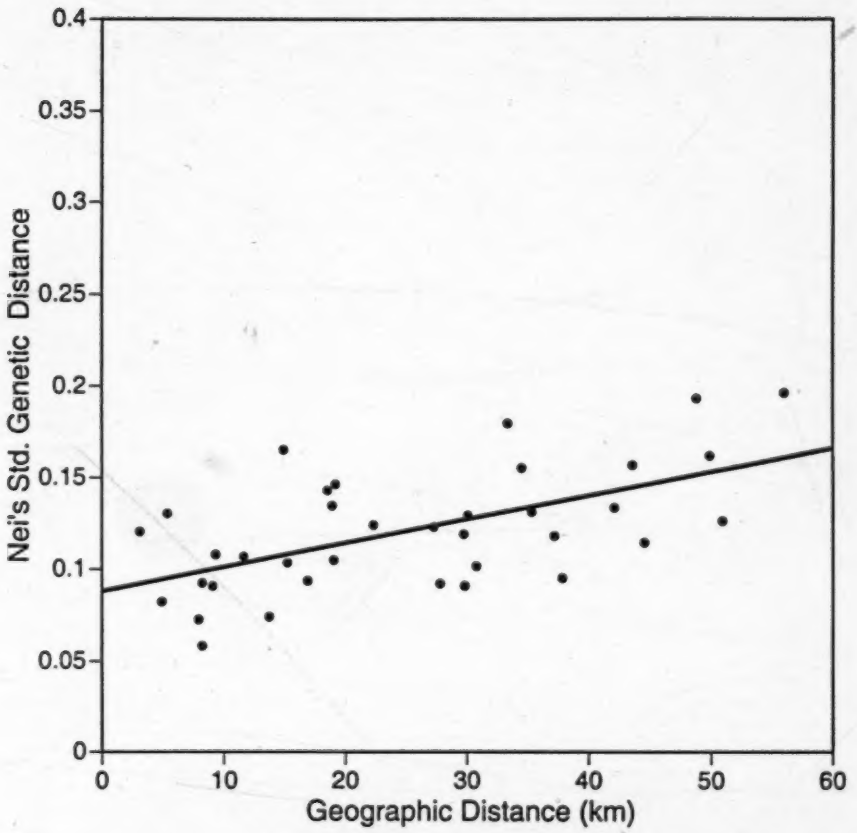


Figure 4-2. Isolation by distance in West Kananaskis: Nei's standard genetic distance versus geographic distance. Each point represents a pair of populations.

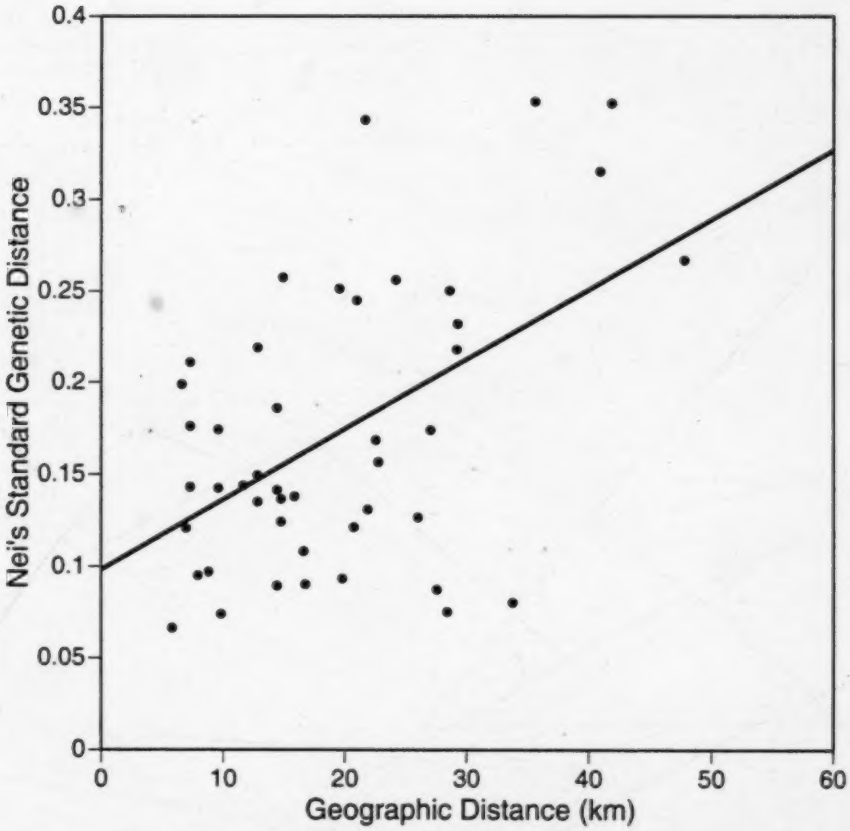


Figure 4-3. Isolation by distance in Banff: Nei's standard genetic distance versus geographic distance. Each point represents a pair of populations.

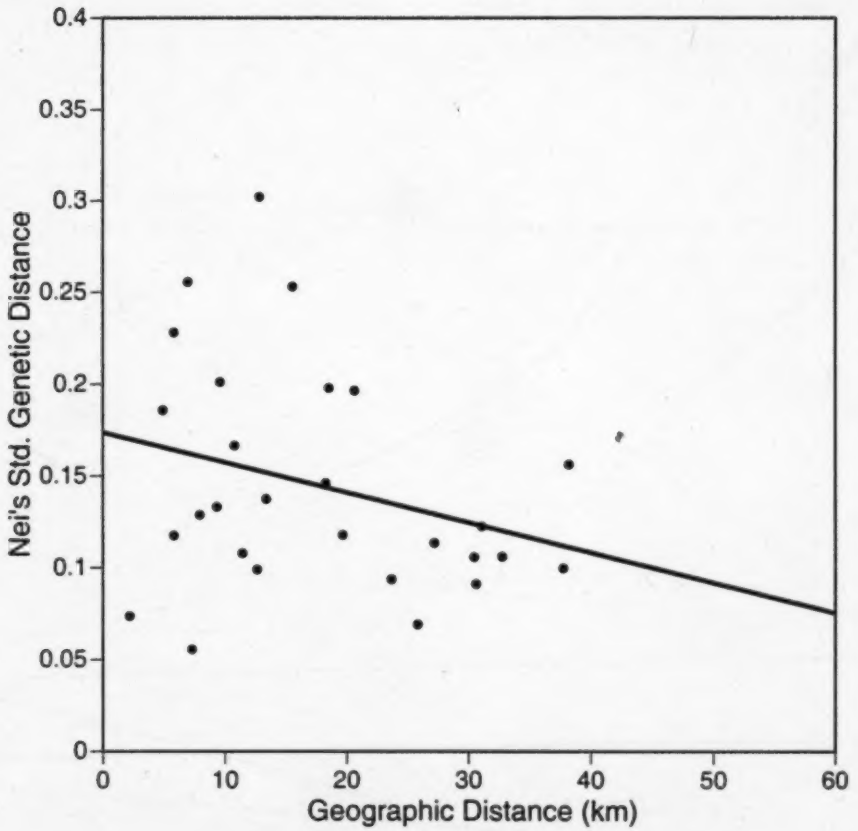


Figure 4-4. Isolation by distance in East Kananaskis: Nei's standard genetic distance versus geographic distance. Each point represents a pair of populations.

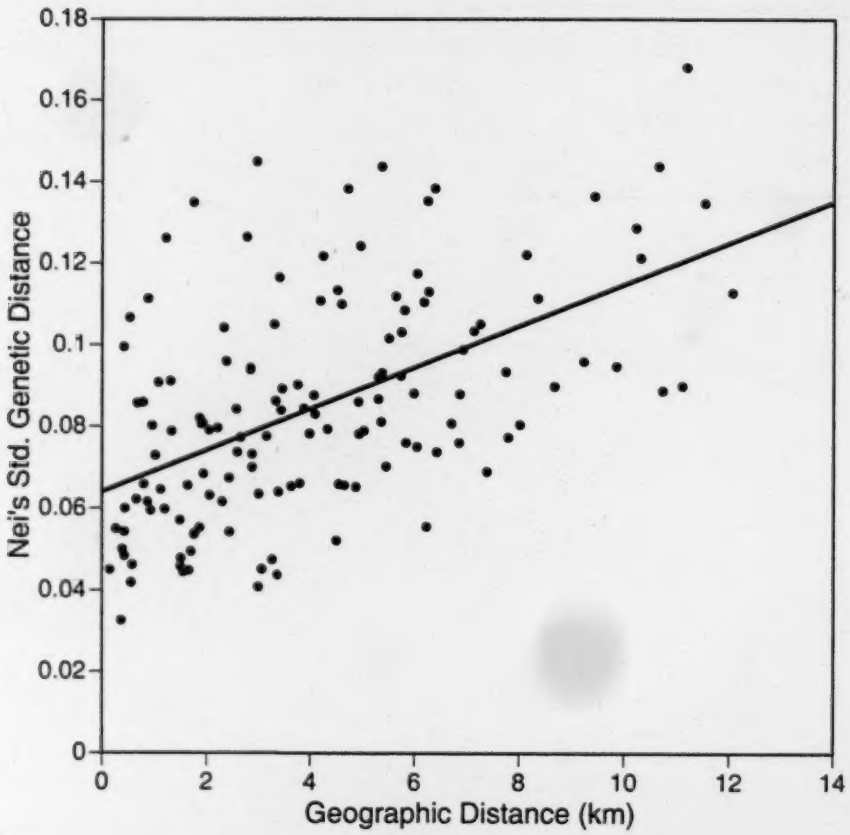


Figure 4-5. Isolation by distance in the small-scale study area (located within East Kananaskis): Nei's standard genetic distance versus geographic distance. Each point represents a pair of populations.

Chapter 5:

General Conclusion

Landscape and population genetic structure

Understanding the effects of landscape structure on movement of individuals and their genes is currently of significant research interest (Wiens 1997, Sork *et al.* 1999, Debinski *et al.* 2001). Incorporation of the effects of landscape structure on movement can add greater realism to theoretical models in ecology and population genetics. Furthermore, a deeper understanding of such effects should allow for better conservation and management of natural populations in the face of rapid and severe landscape changes wrought by human activity.

The results of Chapters 2 - 4 present a fairly consistent picture of the influence of landscape on gene flow and population genetic structure in *Parnassius smintheus*. Population genetic structure is strongly influenced by the surrounding landscape, particularly the relative abundance and distribution of forested versus open, non-forested habitats. This is due to the differential permeability of these different landscape elements to movement of individuals and their genes.

At a fine spatial scale, significant effects of landscape on population genetic structure were observed, paralleling the effects of landscape on movement of individuals (Chapter 2). In particular, I observed greater genetic distances between local populations separated by a given distance of forested habitat, compared to an equivalent distance of non-forested habitat (Chapter 2). This reflected the reduced rates of movement through forested areas that have been observed using mark-release-recapture (MRR) methods (Roland *et al.* 2000). Landscape structure has implications not only for genetic differentiation among sites, but also for levels of genetic variation within sites. In particular, more isolated local populations (based on distances to and sizes of neighbouring local populations) tended to display lower genetic variability than did less isolated local

populations (Chapter 3). However, the effects of surrounding forested and non-forested habitats on levels of genetic variation within sites were not observed, perhaps due to a slow response of within-site heterozygosity to recent habitat fragmentation (Chapter 3). Finally, the effects of landscape on dispersal and genetic structure that were observed at a small spatial scale (up to 12 km separating sample sites) were also apparent at a larger spatial scale (up to 58 km separating sites). A comparison of population genetic structure in three different geographic regions indicated a positive relationship between inferred levels of gene flow, sample heterozygosity, and the connectivity of high altitude, non-forested areas (Chapter 4).

Microsatellites for analysis of population structure in butterflies

Butterflies have served as model systems for the study of population subdivision, spatial structure, and movement of genes and individuals. Studies on the butterflies *Polyommatus icarus* and *Maniola jurtina* by Ford and colleagues in the 1940s represent some of the earliest quantitative investigations of subdivided populations (reviewed in Ford 1964). Studies of *Euphydryas editha* by Ehrlich and co-workers have contributed significantly to our understanding of both ecological and evolutionary processes in subdivided populations (e.g., Ehrlich *et al.* 1975, McKechnie *et al.* 1975). Also, the study of extinction-colonization dynamics in metapopulations has been advanced considerably by the work of Hanski and colleagues on the butterfly *Melitaea cinxia* (e.g., Hanski 1994, Saccheri *et al.* 1998).

This thesis represents some of the first substantial studies of population structure in a butterfly species using microsatellite DNA loci as genetic markers. Despite difficulties with marker development and the occurrence of null alleles at all loci (Chapters 2 & 6), microsatellite allele frequencies proved to be highly informative and revealed genetic structure even at a very fine spatial scale. In particular, they permitted analysis of population genetic structure at the spatial scale of a mark-release-recapture study. This

allowed me to integrate the genetic data with data on individual movements to an extent that previously has not been possible in butterflies, nor in many other organisms (Chapter 2).

Estimation of movement rates at a large spatial scale

At the outset of this project, one of my aims was to estimate rates of movement at a large spatial scale using genetic data. Because mark-release-recapture (MRR) methods are very labour-intensive, they can only be applied at relatively small spatial scales. For example, the study of Roland *et al.* (2000), in which there were only 12 km separating the most distant sites, represents a fairly large scale for a butterfly MRR study. Estimation of movement at larger spatial scales, where MRR is not applicable, usually requires an indirect approach based on inference from observed geographic patterns of genetic variation. By far, the most commonly used approach is to calculate gene flow (the product 'Nm' where N is effective population size and m is the migration rate) from estimates of F_{st} using the equation $F_{st} = 1 / (1 + 4 Nm)$ from Wright's island model of migration. This approach has been criticized however, on both conceptual and statistical grounds (Whitlock & McCauley 1999), and my results suggest that such estimates of gene flow based on microsatellite allele frequencies may be spuriously high because of the high mutation rates and variability of the microsatellites (Chapter 4). A different approach is required for estimating large-scale rates of movement in *P. smintheus*. One possibility is to extrapolate from the relationship between genetic distance and predicted rates of inter-patch movement observed at a fine spatial scale (Fig. 2-3) to estimate rates of movement at a larger spatial scale using observed values of genetic distance.

Figure 5-1 shows the reduced major axis (RMA) linear regression of Nei's standard genetic distance (D_S; Nei 1972) on the average proportion of marked butterflies predicted to move between sites (ln transformed), for all pairs of sites in the fine-scale study area. D_S was based on allele frequencies at seven microsatellite loci (Ps50, Ps76, Ps81, Ps85, Ps163, Ps165, and Ps262). Average proportion of marked butterflies predicted to move

between pairs of sites was calculated as described in Chapter 2 (Fig. 2-3). Using the values of D_S observed in each region in the large-scale study area (Chapter 4), can estimates of movement be obtained from the regression presented in Fig. 5-1?

The proposed approach for estimating movement rates assumes that differences in values of D_S among pairs of populations reflect differences in rates of gene flow. Lack of a pattern of isolation by distance in East Kananaskis was interpreted as reflecting the isolation of populations and the dominant role of genetic drift in influencing allele frequencies and genetic distances (Chapter 4). Therefore, within this region, differences in values of D_S among pairs of populations are primarily not a reflection of differences in levels of gene flow and cannot be used to make inferences about movement. In Banff and West Kananaskis however, a significant pattern of isolation by distance was observed (Chapter 4). Here, D_S is more likely to reflect differences in levels of gene flow and may be used to make inferences about movement. In Banff the mean observed value of D_S was 0.17 with minimum and maximum values of 0.066 and 0.35 respectively. In West Kananaskis the mean observed value of D_S was 0.12 with minimum and maximum values of 0.058 and 0.20 respectively. Inverse prediction from the regression in Fig. 5-1 yields a mean estimate of inter-population movement rates of 0.76×10^{-5}, with maximum and minimum values respectively of 0.008 and 0.44×10^{-10} in Banff. Similarly, in West Kananaskis, the mean estimate of inter-population movement rates is 0.0002, with maximum and minimum values respectively of 0.014 and 0.15×10^{-5}. For comparison, the mean rate of inter-patch movement in the fine-scale study area was 0.0022 with maximum and minimum values respectively of 0.13 and 0.0001. On average, movement rates appear to be approximately 25 times higher in West Kananaskis than in Banff. The maximum predicted movement rates at the large scale appear to be similar to movement rates in the fine-scale study area. Furthermore, the minimum movement rates at the large scale are extremely low and indicate that at larger distances populations are probably completely isolated from each other.

Overall, these estimates of large-scale movement are very approximate. Although the relationship between movement and D_S is highly significant ($P = 0.0003$ from Mantel test), it is 'noisy' with an r^2 value of only 0.25. In addition, estimation of movement at large distances requires extrapolation far beyond the data. Thus, any predictions based on the relationship between movement and D_S are associated with a very large error. This approach does however represent a novel method for obtaining 'ball park' estimates of rates of movement at a large spatial scale.

Alternative methods of genetic data analysis

All of the results presented in this thesis are based on analyses of allele frequencies at microsatellite loci. Other approaches to analysis of microsatellite-based data may add new perspectives and novel insights. Due to their high variability, genotypes at several microsatellite loci can often distinguish individuals and identify familial relationships among individuals. Assignment-tests use genotypic data to describe relatedness of sampled populations and to estimate the sources of migrant individuals (e.g., Paetkau *et al.* 1995). These techniques may allow for even more intimate integration of genetic data with movement data obtained using MRR. Furthermore, given that I have available both detailed movement data from MRR and microsatellite data from the same set of samples, I have a unique opportunity to test some of the limitations of assignment tests. However, any such analyses will require adjustment to account for the occurrence of null alleles at all microsatellite loci in *P. smintheus* (Chapter 2 & 6).

Another approach that would offer a very different perspective of population genetic structure in *P. smintheus* is the estimation of intraspecific phylogenetic relationships using sequence data from one or more independent genetic loci. The use of genealogical information would allow for a coalescent-based approach to data analysis (Hudson 1990) and would also provide a framework to test hypotheses about the relative influences of population history and current gene flow on population genetic structure (Templeton *et al.*

1995, Templeton 1998). Given the high variability observed in the regions flanking microsatellite repeats in this species, it would be particularly interesting to estimate intraspecific phylogenetic relationships and conduct phylogeographic analyses using microsatellite flanking sequences (Chapter 6).

Non-neutral variation

Here, I have examined the effects of landscape structure and dispersal on selectively neutral genetic variation and I have been able to make inferences about the relative strengths of gene flow and genetic drift. To fully understand the evolutionary implications of landscape structure however, it will also be necessary to determine how the forces of drift and gene flow interact with selection pressures to determine patterns of genetic variation in non-neutral, heritable traits. This will require examination of patterns of variation in phenotypic traits of a morphological, biochemical, or behavioural nature that are of consequence to fitness.

Habitat of *Parnassius smintheus*

My analysis of landscape structure has focused largely on the relative abundance and distribution of forested and non-forested habitats at high elevations. This is a relatively simple approach that does not distinguish between the habitat of *P. smintheus*, in which breeding can occur and populations can persist, and other open, non-forested areas in which populations do not persist. For example, in Chapters 2 and 3, the landscape element 'meadow' included both areas in which *P. smintheus* was breeding and areas that did not support the host plant. Similarly, in Chapter 4, I based my measurements of landscape connectivity on the abundance and distribution of all non-forested areas above 2000 m elevation, some of which would not support persistent populations of *P. smintheus*. Therefore, the effects of *P. smintheus* habitat and other types of high-altitude, non-forested

landscape elements were confounded in these studies. This simplification was necessary because defining 'habitat' for *P. smintheus* is not straightforward and many areas containing the host plant do not support populations of *P. smintheus* (Fownes 1999). Thus, determination of habitat and non-habitat for this species is not readily performed using aerial photographs or bio-physical maps, and requires considerable field surveying. However, having established effects of landscape on population genetic structure and gene flow using this relatively simple approach, it would now be informative to try to tease apart the separate effects of *P. smintheus* habitat and other high elevation, non-forested landscape elements. This will likely require Geographic Information System based analyses combined with a rigorous definition of habitat quality for *P. smintheus*, as provided by Fownes (1999), and verification of *P. smintheus* presence/absence with extensive ground truthing.

Butterflies and climate change

There is significant interest in understanding the potential impacts of global climatic warming on biological systems, from a physiological level to an ecosystem level (Peters & Darling 1985, Easterling *et al.* 2000). Research on butterfly species distributions in North America and Europe has provided strong evidence for changes in geographic ranges in response to climate change (Parmesan 1996, Parmesan *et al.* 1999). Here, I consider the potential effects of climate change on natural populations from a different perspective. Rather than focusing on direct effects on the organism of interest, or on changes in the latitudinal distribution of the species' habitat, I consider changes in landscape structure that may occur in response to warming temperatures and that can exert effects on ecological and evolutionary dynamics. My results and those of Roland *et al.* (2000) clearly indicate that changes in the structure of alpine landscapes, via rises in the elevational limit of tree-line, may have very negative consequences for *P. smintheus*. Overall, the impacts of decreasing connectivity of high-altitude meadows brought about by climatic warming are potentially

quite strong and warrant concern and further study. However, the response of *P. smintheus* to climate change will likely be complex and will be affected by more than a rise in tree-line. There may be direct effects of warmer temperatures of butterfly development, physiology, and behaviour (Dennis 1993). There may be changes in the quality of habitat patches that persist. Furthermore, new habitat patches could be created at different altitudes or latitudes. For example, more populations may become capable of occupying lower elevation sites as is currently observed in parts of southern Alberta (Sperling & Kondla 1991). Further complexity is introduced with the possibility of adaptation to changing environmental conditions. Therefore, changes in landscape structure represent only one part of the equation of climate change. Research on many other aspects of warming temperatures are required to predict the potential effects of climate change on this species and to determine what, if any, management techniques could contribute to its preservation.

In summary, I have shown that the landscape within which populations occur can exert a significant influence of patterns of genetic variation in the butterfly *P. smintheus*. My results demonstrate the extent to which interpretation of observed patterns of genetic variation is facilitated by examination of more than one spatial scale and by integration with detailed ecological data. My results also demonstrate the power of microsatellite DNA markers to resolve genetic relationships at very fine spatial scales, such that population genetic and ecological data can be compared directly. I hope that the results presented here, and the approaches I have taken, will contribute to further our understanding of the movements of individuals and their genes over heterogeneous landscapes, and of the role that landscape structure plays in determining ecological and evolutionary dynamics of natural populations.

Literature Cited

Debinski D. M., Ray C., and Saveraid E. H. (2001) Species diversity and the scale of the landscape mosaic: do scales of movement and patch size affect diversity? *Biological Conservation* **98**: 179-190.

Dennis R. L. H. (1993) Butterflies and Climate Change. Manchester University Press, Manchester.

Easterling D. R., Meehl G. A., Parmesan C., Changnon, S. A., Karl T. R. and Mearns L. O. (2000) Climate extremes: observations, modeling, and impacts. *Science* **289**: 2068-2074.

Ehrlich P. R., White R. R., Singer M. C., McKechnie S. W. and Gilbert L. E. (1975) Checkerspot butterflies: a historical perspective. *Science* **188**: 221-228.

Ford E. B. (1964) Ecological Genetics. Methuen & Co. Ltd., London.

Fownes S. L. (1999) Meadow suitability and quality for the alpine butterfly, *Parnassius smintheus*, in the east slopes of the Canadian Rockies. M.Sc. Thesis. University of Alberta.

Hanski I. (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology* **63**: 151-162.

Hudson R. R. (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**: 1-44.

McKechnie S. W., Ehrlich P. R. and White R. R. (1975) Population genetics of *Euphydryas* butterflies I. Genetic variation and the neutrality hypothesis. *Genetics* **81**: 571-594.

Nei, M. 1972. Genetic distance between populations. *American Naturalist* **106**: 283-292.

Paetkau D., Calvert W., Stirling I. and Strobeck C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**: 347-354.

Parmesan C. (1996) Climate and species' range. *Nature* **382**: 765-766.

Parmesan C., Ryrholm N., Stefanescu C., Hill J. K., Thomas C. D., Descimon H., Huntley B., Kaila L., Kullberg J., Tammaru T., Tennent W. J., Thomas J. A., and Warren M. (1999) Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**: 579-583.

Peters R. L. and Darling J. D. S. (1985) The greenhouse effect and nature reserves. *Bioscience* **35**: 707-717.

Roland J., Keyghobadi N. and Fownes S. (2000) Alpine *Parnassius* butterfly dispersal: effects of landscape and population size. *Ecology*.**81**: 1642-1653.

Saccheri I., Kuussaari M., Kankare M., Vikman P., Fortelius W. and Hanski I. (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**: 491-494.

Sork V. L., Nason J., Campbell D. R. and Fernandez J. F. (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution* **14**: 219-224.

Sperling F. A. H.and Kondla N. G. (1991) Alberta Swallowtails and Parnassians: Natural history, keys, and distribution. *Blue Jay* **49**:183-192.

Templeton A. R. (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**: 381-397.

Templeton A. R., Routman E. and Phillips C. A. (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**: 767-782.

Whitlock, M. C. and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. *Heredity* **82**: 117-125.

Wiens J. A. (1997) Metapopulation dynamics and landscape ecology. pp. 43-62 *In* Hanski I. A. and Gilpin M. E. (eds.) Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, San Diego.

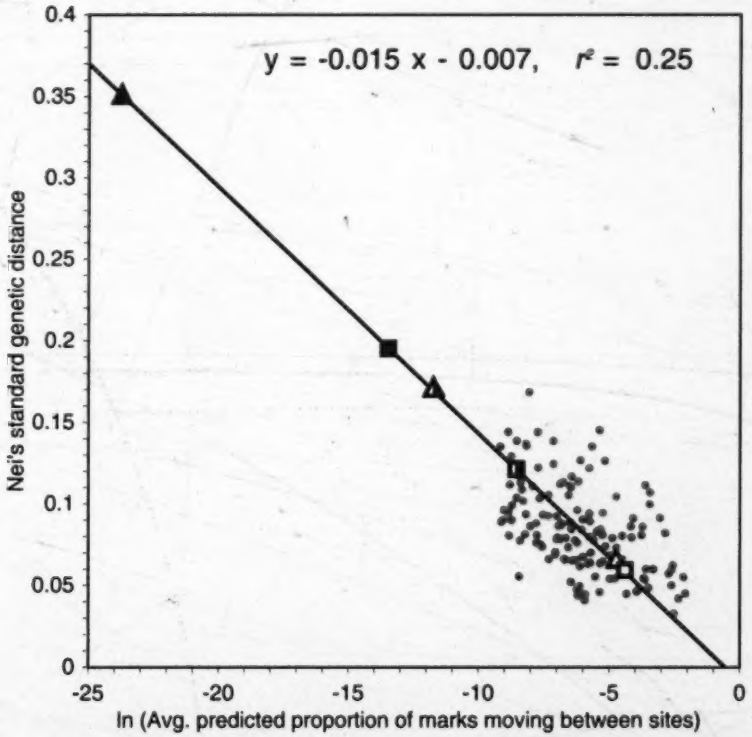


Figure 5-1. Estimation of movement rates at a large spatial scale. Nei's standard genetic distance (based on seven microsatellite loci) versus average proportion of marked butterflies predicted to move between pairs of sites. Grey dots are data from fine-scale study area. Solid line is from reduced major axis regression analysis of fine-scale data. Predicted movement rates (from inverse prediction) in Banff (triangles) and West Kananaskis (squares) are shown, based on minimum (open symbol), mean (half-solid symbol), and maximum (solid symbol) observed values of Nei's genetic distance.

Chapter 6 (Addendum):

Microsatellite loci in a butterfly: flanking sequence variability and other impediments to marker development

Introduction

In less than a decade, microsatellite DNA loci have emerged as arguably the most popular molecular markers for ecological and evolutionary studies. They are prized for their high variability, co-dominant Mendelian inheritance, and broad distribution throughout the genome. They have been widely used to study population genetic structure, kinship, and paternity. Butterflies have served as model systems in ecological and evolutionary studies, in part because they are often intimately associated with specific habitats, they are easily monitored and handled, and knowledge of their natural history is extensive (Boggs *et al.*, *in press*). Given the utility of butterflies as model systems in ecology and evolution, and the value of microsatellite markers as analytical tools in these fields, the application of microsatellite markers to studies of butterfly biology would seem very promising. Unfortunately, progress in this area has been hampered by the difficulty of isolating useful microsatellite markers from butterfly genomes (Meglécz & Solignac, 1998; Keyghobadi *et al.*, 1999; Nève & Meglécz, 2000). Very few microsatellite loci have been reported for butterflies or other Lepidoptera (Palo *et al.*, 1995; Bogdanowicz *et al.*, 1997; Meglécz & Solignac, 1998; Keyghobadi *et al.*, 1999; Reddy *et al.*, 1999; Harper *et al.*, 2000) and isolation and characterization of microsatellite markers is clearly more difficult in these insects than in most other organisms (Meglécz & Solignac, 1998; Nève & Meglécz, 2000).

Isolation of useful microsatellite markers from a given taxon may be difficult for two possible reasons. First, microsatellite sequences may be rare. Relative abundances of different types of microsatellite sequences are known to vary significantly among taxa, and some repeat sequences are far less common in the genomes of certain insects than in

vertebrates (Navajas *et al.*, 1998; Day & Ready, 1999). Second, the microsatellites and their surrounding regions may display characteristics that either impede successful polymerase chain reaction (PCR) amplification or, when amplified, produce banding patterns that are difficult to interpret. Microsatellites are also known to display structural differences and behave differently in different genomes. For example, Brooker *et al.* (1994) found that microsatellites were longer and more likely to be imperfect in teleost fishes as compared to mammals. Even within the same species, different classes of microsatellite repeats can behave quite differently; Bell & Ecker (1994) found that $(CA)_n$ and $(GA)_n$ loci in the plant *Arabidopsis thaliana* differed substantially in structure, levels of polymorphism, and refractoriness to PCR.

Microsatellite sequences do appear to be relatively rare in butterflies (Meglécz & Solignac, 1998; Nève & Meglécz, 2000), but this alone should not pose a serious problem given the wide array of microsatellite enrichment protocols that are now available (e.g., Kijas *et al.*, 1994; Paetkau, 1999). A more serious problem is posed by structural differences in microsatellites or in their flanking sequences that make it difficult to amplify loci or interpret the observed banding patterns (Meglécz & Solignac, 1998). We have previously suggested (Keyghobadi *et al.*, 1999) that in butterflies the regions flanking microsatellite repeats are highly variable, leading to difficulties with primer design and a high frequency of null alleles. This hypothesis was based on the behaviour of four loci in the butterfly *Parnassius smintheus* Doubleday (1847) and seemed to be corroborated by the few other existing reports of butterfly microsatellites. Here, we describe four novel microsatellite loci useful for population studies in *P. smintheus*. We review all the loci that we have attempted to amplify in this species to determine what proportion could be cleanly amplified and easily interpreted, and examine the most common problems with amplification and interpretation. Finally, to test the hypothesis of high variability in the flanking regions, we present sequences of four alleles of each of four loci.

Results

Isolation of novel microsatellite markers

Four novel microsatellite markers useful for genotyping in *P. smintheus* were identified with the microsatellite isolation protocol described by Grist *et al.* (1993). This protocol isolates microsatellite sequences from a small-insert genomic library using the polymerase chain reaction (PCR): single molecules containing the microsatellite sequence of interest are amplified in reactions using a primer designed to recognize the tandem repeat. Once a single microsatellite-containing region is amplified, it can be sequenced to determine the flanking regions and locus-specific primers can be designed. Three-hundred and twenty reactions, or "working stocks" (Grist *et al.*, 1993), were screened for $(GT)_n$ repeats and 23 putative loci were identified. Of the 23 microsatellites identified, we could obtain sufficient flanking sequence on both sides to design locus-specific primers for nine of them. One of these nine was identical to a microsatellite clone that had previously been characterized (Ps162 from Keyghobadi *et al.* (1999)). Thus, we designed PCR primers for eight novel loci. Four of these amplified cleanly and were polymorphic, and were therefore considered useful for genotyping individuals and for population studies. Primer sequences for these four loci are presented in Table 6-1. Redesigning primers or altering PCR stringency did not improve amplification for the other loci (Table 6-2).

The original primers designed for locus Ps76 seemed to produce many null alleles, as most typed individuals appeared homozygous even though many different alleles were present. Redesigning the unlabeled primer (Table 6-1) for this locus reduced the frequency of the null allele and many individuals originally appearing homozygous were heterozygous when retyped.

Locus Ps165 had a very wide range of allele sizes from 180bp to 350bp. The larger alleles at this locus, greater than 260bp in length, displayed many stutter bands, all of

approximately equal intensity, making it difficult to identify the true allele. Thus, for this locus, all alleles greater than 250bp were pooled into a single size class.

Variability, Hardy-Weinberg equilibrium, and null alleles

The four novel microsatellite loci were used to genotype 557 individuals from 17 sampling sites in the Kananaskis region of Alberta, Canada. These individuals have previously been typed at four other loci (Keyghobadi *et al.*, 1999). Generally, the novel microsatellite markers were highly variable. Loci Ps76, Ps262, and Ps165 were particularly variable having 29, 33, and over 36 alleles, and mean expected heterozygosities of 0.84, 0.92, and 0.83, respectively (Table 6-1).

Exact tests for Hardy-Weinberg equilibrium were conducted for each new locus in each of the 17 sites with the program GENEPOP, version 3.1d (Raymond & Rousset, 1995). Significant homozygote excess was observed at 16 sites for Ps76, at all sites for locus Ps165, and at seven sites for Ps262. With a Bonferroni adjustment of the α-value to correct for total experimentwise error ($\alpha = .05/68 = .0007$), a total of 31 tests at loci Ps76, Ps165, and Ps262 were still significant. No significant deviations from Hardy-Weinberg equilibrium were observed for locus Ps163. However, this probably reflects a lack of power to detect deviations from expected proportions at this locus because at all sites a single allele appeared at very high frequency (mean frequency = 0.88) and most individuals were homozygous for that allele.

We observed 10 apparent null homozygotes at locus Ps76, six at Ps163, and 66 at Ps165. For these individuals, no PCR product was obtained at that locus, even with repeated attempts, though genotypes were readily obtained at other loci. The occurrence of such individuals strongly suggests that these loci suffer from null alleles.

Homozygote excess at locus Ps262 in seven populations suggests this locus also displays null alleles. The presence of null alleles at this locus is more equivocal however, as no null homozygotes were observed. In addition, we could not use the band intensities

from electropherogram data to confirm the presence of null alleles at this locus as we have previously done for Ps50 (Keyghobadi *et al.*, 1999); the large number of alleles and high heterozygosity at Ps262 resulted in few individuals of the same genotypes whose band intensities could be compared. However, we suggest that null alleles occur at this locus as well, as other explanations for excess homozygosity are not likely (see Discussion).

For each locus in each population, an estimate of the frequency of the null allele was obtained using the estimation-maximization (EM) algorithm as described in Keyghobadi *et al.* (1999). Mean null allele frequency varied widely among loci and was as high as 0.36 at locus Ps165 (Table 6-1). For all four loci, there was no significant difference between the observed number of null homozygotes within sites and the expected number given the estimated frequency of the null allele (Ps76: $\chi^2_{[df=16]} = 22.9$, $P = 0.12$; Ps 163: $\chi^2_{[df=12]} = 3.4$, $P > 0.99$; Ps165: $\chi^2_{[df=16]} = 7.8$, $P = 0.95$; Ps262: $\chi^2_{[df=12]} = 1.16$, $P > 0.99$).

P. smintheus *microsatellites: sequence, length, and utility for genotyping individuals*

Table 6-2 lists all microsatellite loci isolated from *P. smintheus* for which PCR primers have been designed. Isolation of these loci is described either here or in Keyghobadi *et al.* (1999) and was performed using either the PCR-based protocol (Grist *et al.*, 1993) or a standard protocol involving bacterial transformation and plaque screening (e.g., Paetkau & Strobeck, 1994). Screening was always for $(GT)_n$ repeats.

Weber's (1990) criteria were used, with slight modification, to delimit microsatellite sequences and categorize them as perfect, imperfect, or compound; because we have found that in this species even short repeats can be quite variable, we classified as few as four tandem repeats as a perfect sequence and for compound sequences we required only three tandem repeats of the second motif. Thus, we observed 14 perfect (70.0%), 5 compound (25.0%), and 1 imperfect (5.0%) microsatellites. Loci Ps92 and Ps248 each contained two perfect microsatellites separated by 28 bp and 8 bp of intervening sequence, respectively.

These were counted separately. The average length of the longest uninterrupted run of repeats at each locus was 10.3 (± 1.5 SE) repeat units, and differed slightly between isolation protocols. Average, uninterrupted repeat length was 13.0 (± 4.3 SE) units for loci isolated with the standard protocol and 9.0 (± 0.9 SE) units for loci isolated with the PCR-based method. This difference was not significant ($t = 1.24$, $df = 16$, $P = 0.23$) and was largely due to a single very long repeat (Ps67) isolated with the standard protocol.

Loci judged "usable" for genotyping amplified cleanly with only one or two bands in each individual, displayed allelic variation, and yielded products in most (*i.e.* in >90% of) individuals. Of 18 independent loci, only seven were found to be usable (Table 6-2). Of the 11 that were not usable, seven showed multiple banding with more than two bands per individual. In all cases, varying PCR stringency or re-designing primers did not eliminate the multiple banding pattern (Table 6-2). Locus Ps162 was originally found to be useful for genotyping populations in Kananaskis, Alberta, Canada (Keyghobadi *et al.*, 1999) but was later found to display multiple banding in several populations from Banff National Park, Alberta, Canada. Two loci yielded no product of the expected size, one was invariable, and one (Ps135) suffered from such a high frequency of null alleles that no product was observed in many individuals. In addition, all seven usable loci appeared to have null alleles at a high enough frequency to cause detectable homozygote excess or to produce null homozygote individuals. Altering PCR stringency (Table 6-2) did not reduce the incidence of null alleles at any loci. For several loci however (Ps81, Ps85, Ps76), redesigning primers did reduce the number of null alleles, but never eliminated them (Table 6-2; Keyghobadi *et al.* 1999) Thus, the most common amplification patterns for *P. smintheus* loci overall were multiple banding and null alleles. For all loci, all primers tested amplified single bands of the expected size when the original phage stock or working stock was used as template.

Sequencing of microsatellites and flanking regions

We sequenced four alleles representing two-size classes for each of loci Ps76, Ps81, Ps165, and Ps262 (Fig. 6-1). In all cases allelic size variation was associated with different numbers of repeats in the microsatellite region. However, these sequences also reveal a high frequency of mutation in the microsatellite flanking regions.

Locus Ps76 had 213 bp of flanking sequence not including the primer sequences and two repeat regions, one of which was not recognized as a potentially variable site in the original clone where it was $(AT)_3$. Within the 213 bp of flanking sequence we observed 12 base substitutions (nine transitions and three transversions) and one single-bp indel. Locus Ps81 had 76 bp of flanking sequence with one transitional substitution. Locus Ps165 had 141 bp of flanking sequence with one transitional substitution and one single-bp indel. One transitional substitution was also observed in the microsatellite repeat of Ps165. Locus Ps262 had 43bp of flanking sequence with four base substitutions (one transition and three transversions). The microsatellite repeat in Ps262 was also interrupted by a single transversion.

Base substitutions in the flanking regions were highly correlated within loci, and we could identify "families" of alleles sharing a particular flanking sequence haplotype at each locus (Fig. 6-1). We observed two cases of size homoplasy. At locus Ps76, the 284 bp alleles of Forget-Me-Not 30 and Flint 22 differed in the number of (GTT) and (AT) repeats, by a single-bp indel, and by several base substitutions. At locus Ps262, the 102 bp alleles of Mt. Kidd 7 and REB differed by several base substitutions. Single-bp size variation among alleles was caused by both indels in the flanking regions (loci Ps76 and Ps165) and variation in tandem repeats of three-bp microsatellite motifs associated with the two-bp motif (locus Ps76).

Among the four loci, we did not observe a positive relationship between the proportion of flanking sites that were variable and the average null allele frequency. For example, more flanking sequence variability was observed at locus Ps262 which has a

relatively low average null allele frequency (3.2%) than at locus Ps165 which has the highest average null allele frequency (35.8%).

Discussion

Microsatellite repeats appear to be relatively rare in butterflies as compared to many other species (Meglécz & Solignac, 1998; Nève & Meglécz, 2000). Our results reveal that microsatellites and their flanking regions may also behave quite differently in butterflies than in other taxa. In the butterfly *P. smintheus*, PCR amplification of microsatellite loci often produced multiple bands that were difficult to interpret. Loci that amplified cleanly displayed a high frequency of null alleles and exhibited high variability in the flanking regions. Thus, it is probably the combination of low frequency and unusual characteristics that has hampered development of microsatellite markers in this well-studied group of insects.

Amplification of microsatellite loci

A high proportion of microsatellite loci from *P. smintheus* yielded multiple bands upon amplification (Table 6-2). Varying PCR reagent concentrations and cycling conditions did not eliminate the multiple bands. As single-locus and co-dominant inheritance is a major strength of microsatellites, loci displaying multiple banding patterns are not considered useful molecular markers. Thus, a low proportion of the loci isolated from *P. smintheus* were useful. A similar situation exists in some plants with complex genomes, where a very low proportion of isolated microsatellite sequences can be amplified successfully and this is considered a major impediment to development of microsatellite markers in these species (Röder *et al.*, 1995; Ma *et al.*, 1996; Pfeiffer *et al.*, 1997; Buteler *et al.*, 1999). Röder *et al.* (1995) also found that adjustment of PCR conditions did not improve amplification for primer pairs that were originally problematic.

Multiple banding could be due to the occurrence of microsatellites within larger repeated elements. We have observed this phenomenon in one microsatellite-containing clone from *P. smintheus* (Ps09; Accession # AY048082; isolated with bacterial transformation), where the cloned fragment of *P. smintheus* DNA contained two tandem repeats of a 224 bp long sequence containing a $(GT)_4$ repeat. Location of microsatellite sequences in larger repeated elements has been observed in other organisms. In humans, Weber (1990) noted that $(GT)_n$ repeats are known to occur within introns of the two γ-globin genes and within a portion of Alu elements, and that 7% of primer pairs for $(GT)_n$ repeats tested in humans have yielded three or more alleles per individual. Multiple banding has also observed at a large proportion of microsatellite loci in coniferous tree species and is due to their occurrence within highly repetitive DNA (Smith & Devey, 1994; Kostia *et al.*, 1995; Pfeiffer *et al.*, 1997; Karhu *et al.*, 2000). Specific amplification patterns for unusable microsatellite loci have not been reported for other butterfly species. However, Reddy *et al.* (1999) identified 15 microsatellite loci in the silkworm moth, *Bombyx mori,* at least some of which appear to show multiple-banding patterns upon PCR amplification (e.g., locus *sat211*) and are interpreted as representing duplicated loci. Overall, further data on the sequences of the multiple bands and on the abundance of repetitive DNA in the genome are required to confirm the source of the multiple banding patterns in *P. smintheus*.

Homozygote excess and null alleles

The usable microsatellite loci isolated from *P. smintheus* were highly variable (Table 6-1) and could differentiate among populations over a small geographic scale (Keyghobadi *et al.*, 1999). However, they also had a high frequency of null alleles, suggesting variability in the flanking regions.

Genotype frequencies at almost all loci used for genotyping (Ps50, Ps81, Ps85, Ps76, Ps163, Ps165, Ps262) exhibited homozygote excess as compared to Hardy-

Weinberg (HW) expectations in many populations. Usually, deviation from HW equilibrium at all loci is attributed to some population-level process such as the Wahlund effect and not to individual, locus -specific problems such as null alleles. Here, though six of seven loci exhibit heterozygote deficits, we argue that null alleles are the main contributing factor. First, at five of those six loci there is independent evidence for null alleles in the form of null homozygote individuals or from analysis of electropherogram data (Keyghobadi *et al.*, 1999). Second, other possible population-level causes of homozygote excess seem unlikely in these populations. For neutral molecular markers, probable population-level causes of homozygote excess are the Wahlund effect or inbreeding via assortative mating of related individuals. The Wahlund effect is extremely unlikely in this case because we sampled at such a fine spatial scale; we are far more likely to have erred in the opposite direction and to have divided a largely homogeneous population into two or more separate samples. Inbreeding through excessive mating of related individuals in also unlikely. Females lay eggs individually and can travel tens of meters between successive ovipositions (Fownes, 1999). Overwintering occurs in the egg stage and must be associated with a high mortality rate, given that a single female is capable of laying hundreds of eggs (Fownes, 1999). Larvae are not gregarious. Males patrol areas in search of females and newly emerged virgin females tend to be mated quickly upon eclosure. Given these aspects of the life-history and behaviour of *P. smintheus*, there is no reason to suspect that a disproportionate number of matings among related individuals normally occurs.

Allelic drop-out, or the amplification of only one allele in a heterozygous individual due to chance variation in the amount of template, is also not a probable explanation for the observed homozygote excess at these loci. Any samples that were amplified several times yielded very consistent results and decreasing PCR stringency did not reveal new alleles. For loci where a high proportion of homozygous or unamplifiable individuals were initially

observed (e.g., Ps76), only new primers and not adjustment of assay conditions revealed new alleles.

Significant homozygote excess has been noted in all reports of butterfly microsatellites (Palo *et al.*, 1995; Meglécz & Solignac, 1998; Keyghobadi *et al.*, 1999; Harper *et al.*, 2000), though this is not always attributed to null alleles.

Variability of microsatellite flanking regions

Sequencing of alleles confirmed the hypothesis (Keyghobadi *et al.*, 1999) that microsatellite repeats in *P. smintheus* are flanked by highly variable sequences (Fig. 6-1). Analysis of only four alleles of each of four microsatellite loci in *P. smintheus* revealed polymorphism in flanking sequences of all loci. And, flanking sequences of two loci (Ps76 and Ps262) exhibited extremely high variability, with base substitutions at 5.6% and 9.3% of flanking sites, respectively. These values are comparable to interspecific levels of polymorphism observed in microsatellite flanking regions in other groups. In several reports of interspecific variation in microsatellite and flanking region sequences, the average proportion of flanking sites with base substitutions was 8.2 % (range: 0.0 - 16.8%; Schlötterer *et al.*, 1991; Blanquer-Maumont & Crouau-Roy, 1995; Angers & Bernatchez, 1997; Clisson *et al.*, 2000; Karhu *et al.*, 2000; Makova *et al.*, 2000; van Oppen *et al.*, 2000). Higher levels of base substitution in microsatellite flanking regions have been reported (FitzSimmons *et al.*, 1995; Rico *et al.*, 1996; Primmer & Ellegren, 1998), but these all include comparisons of distantly related species in different families or orders. Most intraspecific polymorphism in microsatellite flanking regions involves indels rather than base substitutions (Grimaldi & Crouau-Roy, 1997; Colson & Goldstein, 1999; Anderson *et al.*, 2000). Levels of base substitution in microsatellite flanking sequences within species similar to those we observed are rare and restricted to single loci (Ortí *et al.*, 1997; Primmer & Ellegren, 1998). Thus, for comparison, Makova *et al.* (2000) found base substitutions at a maximum of only 0.5% of microsatellite flanking sites within

species of mice (*Apodemus* sp.). In addition, Viard *et al.* (1998) found no intraspecific variation in microsatellite flanking sequences in the honey bee (*Apis mellifera*), a bumble bee (*Bombus terrestris*), nor in a freshwater snail (*Bulinus truncatus*), despite sequencing over 40 alleles at some loci. Few data have been presented on microsatellite flanking region variability in other butterflies or moths. However, Palo *et al.* (1995) report atypical sequence rearrangements among alleles of two microsatellite loci in the Glanville fritillary, *Melitaea cinxia*. Also, Rohwedel *et al.* (1993) found that very long $(GATA)_n$ repeats (*i.e.* hundreds of repeats long) in the mealmoth *Ephestia kuehniella* tended to be surrounded by unstable sequence.

Variability in the flanking regions of microsatellite loci has several implications for their use as molecular markers. First, primer design can be challenging. For some *P. smintheus* loci, a number of primer combinations were tested before amplification products could be obtained in most individuals. At one locus (Ps135) this was never achieved. Also, despite considerable effort in testing different primer combinations (Table 6-2), null alleles were never eliminated from any loci. Second, simple mutational models such as the stepwise mutation model are inappropriate for these loci as much variation, including significant size variation (Keyghobadi *et al.*, 1999), can occur in the flanking region. Finally, we predict a low success rate with cross-species amplifications in butterflies. Our attempts to amplify *P. smintheus* loci in another *Parnassius* species have been largely unsuccessful. Only one locus (Ps262) could be successfully amplified, and displayed variable and interpretable banding patterns, in a sample of seven individuals of *Parnassius apollo*. This is in sharp contrast to groups in which cross-species amplification has been successful over millions of years of evolution (FitzSimmons *et al.*, 1995; Rico *et al.*, 1996).

As null alleles at microsatellite loci are usually attributed to mutations in the primer-binding sites of the flanking regions (Callen *et al.*, 1993; Paetkau & Strobeck, 1995), we would generally expect a positive correlation between flanking sequence variability and null

allele frequency among loci. We did not observe such a correlation. However, given the small sample size and the fact that the flanking sequence variation did not include the primer binding sites, this result is perhaps not surprising.

Despite the instability of the flanking regions, most of the length variation among alleles was due to variation in the number of tandem repeats, which may arise via strand slippage during replication (Schlötterer & Tautz, 1992). Base substitutions in the flanking regions were highly correlated and we could identify "families" of alleles within loci, members of which presumably share a common ancestry. This suggests that microsatellite flanking regions may provide informative sequences for phylogeographic analysis. By analogy, variation in microsatellite flanking sequences has been used for reconstructing interspecific phylogeny in other taxa (Zardoya *et al.*, 1996).

High variability may be a shared property of flanking regions and tandem repeats of microsatellites. Glenn *et al.* (1996), in testing the amplification of American alligator microsatellites in a range of crocodilian genera, found a negative correlation between the variability of a locus in alligators and the number of crocodilian genera in which it could be amplified. In comparing sequences of three microsatellites and their flanking regions across several bird species, Primmer & Ellegren (1998) found the one locus that displayed the highest level of flanking sequence variation was also extremely variable. Thus, there may generally be an association between high variability in the repeat region and variation in the flanking regions, suggesting microsatellites and their flanks experience similar mutation and/or selection pressures. Three of the four loci for which we sequenced alleles show similar levels of variation so it is difficult to test directly for a correlation between variability in the repeat and flanking regions in these loci. However, our data do support this hypothesis in a more general sense, as microsatellite loci from *P. smintheus* as a whole are characterized by both high variability and high flanking sequence instability.

The high variability of microsatellite loci and their flanking regions in *P. smintheus* may be explained partly by the species' relatively short generation time (one year).

However, other animals with short generation times, such as mice and some Hymenopterans, do not exhibit the same high rates of mutation in microsatellite flanking sequences (Viard *et al.*, 1998; Makova *et al.*, 2000). Other possible explanations for the higher mutation rate in *P. smintheus* are required. However, the factor(s) contributing to the high variability of microsatellites and their flanking regions in this species may be complex, as nucleotide substitution rates can vary both across species and among different parts of the genome within species (e.g., Caccone & Powell, 1990), due to differences in selective constraints (Li & Graur, 1991), mutation rates (Wolfe *et al.*, 1989), or both.

It is becoming increasingly clear that not only the abundance of microsatellite sequences, but also their qualitative characters and the mutational dynamics of the repeat regions and their flanking sequences may differ considerably among genomes. There is clearly some factor other than low abundance that has impeded the development of microsatellites as molecular markers in butterflies and possibly other Lepidoptera. We have presented data from the butterfly *P. smintheus* confirming that microsatellite loci in this species are flanked by highly variable sequences, and we have suggested that the microsatellites may also occur within larger, repeated elements. These results are largely from a single class of microsatellite repeat in a single species. Though observations from several butterfly and moth species tend to support our conclusions, extrapolation to other repeat classes and other species must be made with caution. Analysis of the behaviour of microsatellite repeats and their flanking regions in other Lepidoptera is required to determine the generality of the results presented here.

Experimental Procedures

Microsatellite isolation

Novel microsatellites were isolated using the PCR-based protocol described by Grist *et al.* (1993). Details of the methodology are the same as in Keyghobadi *et al.* (1999). Briefly,

P. smintheus genomic DNA was digested with *Sau*3A1 and fragments 200-600 bp long were ligated into *Bam* H1-digested and dephosphorylated M13mp18 vector. Ten-fold serial dilutions of the products of ligation were used as template in PCR amplifications with a microsatellite-specific primer $(GT)_{11}$ and a modified M13 universal forward sequencing primer (forward: 5'CGACGTTGTAAAACGACGGCCAG3'). Amplifications were in 25 µL with 1X PCR buffer (10 mM Tris, pH 8.8, 0.1% Triton X100, 50 mM KCl and 0.16 mg/ml BSA), 2 mM $MgCl_2$, 0.16 µM of each primer, 120 µM dNTP, 0.5 units of *Taq* DNA polymerase (purified as described by Engelke *et al.* (1990)), and 1 µL of diluted ligation reaction. The cycling conditions were 1 min. at 94 °C, followed by 35 cycles of 30 sec. at 94 °C, 20 sec. at 56 °C, and 7 sec. at 72 °C, followed by 30 sec. at 72 °C, on a 9600 thermal cycler (Perkin-Elmer). Products were run on agarose gels to determine the level of dilution at which discrete bands would be produced. At dilutions lower than this critical level smears would be observed and at higher dilutions no product would be observed. A large number (*i.e.*. hundreds) of individual "working stocks" (Grist *et al.* 1993) were made by using 1, 2, 5, or 10 µL of the critical ligation dilution as template in a similar PCR, but with only 20 cycles and with the M13 universal forward and reverse sequencing primers (reverse: 5'CAGGAAACAGCTATGACC3'). Working stocks were then diluted 100-fold and 1 µL of this dilution was used as template in a PCR with the $(GT)_{11}$ microsatellite-specific primer and the M13 forward sequencing primer under the same amplification conditions as in the first round of PCRs. The products of these reactions were run on agarose gels and individual microsatellite loci were identified as sharp bands. DNA from each such band was purified and sequenced from the forward sequencing primer with a dRhodamine Terminator Cycle Sequencing Kit (Applied Biosystems). Extension products were resolved on a 377 automated sequencer (Applied Biosystems). This provided flanking sequence for one side of the microsatellite. A primer was designed from this sequence and used with the reverse M13 sequencing primer to amplify across the microsatellite. These products were sequenced from the reverse

sequencing primer to obtain the flanking sequence on the other side of the microsatellite. Optimal primer pairs for amplification of the microsatellite region were designed with OLIGO software (National Biosciences Inc., Version 5.0).

Sample collection and DNA extraction

Tissue samples of 557 adult butterflies were collected in 1995 and 1996 from seventeen sites along three ridges in the Kananaskis region of Alberta, Canada as described in Keyghobadi *et al.* (1999). Sampling was at a very fine scale, with only 12 km separating the most distant sampling sites. Genomic DNA was extracted with the QIAamp™ tissue extraction kit (QIAGEN). The samples consisted of either small wing clippings (approximately 0.15 cm^2) or the entire body. In the former case, the entire sample was used for DNA extraction. In the latter case, approximately 25 mg of abdominal or thoracic tissue was used for males and females, respectively.

Microsatellite amplification and analysis

All novel loci were amplified separately in 15μL reactions containing 1X PCR buffer (10 mM Tris, pH 8.8, 0.1% Triton X100, 50 mM KCl and 0.16 mg/ml BSA), 2 mM $MgCl_2$, 0.16 μM of each primer, 120 μM dNTP, 0.3 units of *Taq* DNA polymerase, and approximately 75 ng of genomic DNA. Cycling conditions were 1 min. at 94 °C, followed by three cycles of 30 sec. at 94 °C, 20 sec. at 54 °C, and 5 sec. at 72 °C, followed by 33 cycles of 15 sec. at 94 °C, 20 sec. at 54 °C, and 1 sec. at 72 °C, followed by a 30 min. final extension at 72 °C, on a 9600 thermal cycler (Perkin-Elmer). One primer of each primer pair was end-labeled with either 6-FAM, HEX, or TET fluorescent dye (Applied Biosystems). Products were electrophoresed and detected on a 373A automated sequencer (Applied Biosystems) and electrophoretic data were collected and analyzed with GENESCAN and GENOTYPER software (Applied Biosystems).

Sequencing of microsatellites and flanking regions

For each of four loci (Ps76, Ps81, Ps165, and Ps262), four different alleles representing two size classes were sequenced. Sequences were obtained from some individuals sampled at the study area described here (see *Sample collection and DNA extraction*) , as well as some individuals sampled at other sites in Kananaskis and Banff National Park, Alberta, Canada. For each locus, we identified heterozygous individuals having the allele of the desired size with a difference of at least 20 bp from the other allele. In these individuals, the microsatellite locus of interest was amplified using chimeric primers: M13 forward and reverse sequencing primers were appended to the 5' ends of the original primer sequences (Paetkau & Strobeck, 1995). PCR conditions and cycling parameters were identical to those we normally used for microsatellite amplification, except reactions were in 100μL total volume. Primer sequences and PCR conditions for Ps81 are provided in Keyghobadi *et al.* (1999). As we selected heterozygous individuals with alleles of very different sizes, the two alleles were easily separated by electrophoresis of the PCR products on 1.5% agarose gels. The allele of interest was excised from the gel and the DNA was extracted with a QIAquick™ gel extraction kit (QIAGEN). This approach took advantage of the large range of allele sizes observed at some loci to allow us to unambiguously sequence single alleles without having to clone the PCR products. The products of gel extraction were sequenced in both directions with M13 forward and reverse sequencing primers and a BigDye™ Terminator Cycle Sequencing Kit (Applied Biosystems) on a 377 automated sequencer (Applied Biosystems).

Literature cited

Anderson, T. J. C., Su, X., Roddam, A. and Day, K. P. (2000) Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. *Mol Ecol* **9**: 1599-1608.

Angers, B. and Bernatchez, L. (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Mol Biol Evol* **14**: 230-238.

Bell, C. J. and Ecker, J. R. (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**: 137-144.

Blanquer-Maumont, A. and Crouau-Roy B. (1995) Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. *J Mol Evol* **41**: 492-497.

Bogdanowicz, S. M., Mastro, V. C., Prasher, D. C. and Harrison, R. G. (1997) Microsatellite DNA variation among Asian and North American gypsy moths (Lepidoptera: Lymantriidae). *Ann Ent Soc Amer* **90**: 768-775.

Boggs, C., Watt, W. and Ehrlich, P. (In Press) Ecology and Evolution Taking Flight: Butterflies as Model Study Systems. University of Chicago Press. Chicago.

Brooker, A. L., Cook, D., Bentzen, P., Wright, J. M. and Doyle, R. W. (1994) Organization of microsatellites differs between mammals and cold-water teleost fishes. *Can J Fish Aq Sci* **51**: 1959-1966.

Brownstein, M. J., Carpten, J. D. and Smith, J. R. (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques* **20**: 1004-1010.

Buteler, M. I., Jarret, R. L. and LaBonte, D. R. (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. *Theor App Genetics* **99**: 123-132.

Caccone, A. and Powell, J. R. (1990) Extreme rates and heterogeneity in insect DNA evolution. *J Mol Evol* **30**: 273-280.

Callen, D. F., Thompson, A. D., Shen, Y., Phillips, H., Richards, R. I., Mulley, J. C., and Sutherland, G. R. (1993) Incidence and origin of null alleles in the $(AC)_n$ microsatellite markers. *Am J Hum Genet* **52**: 922-927.

Clisson, I., Lathuilliere, M. and Crouau-Roy, B. (2000) Conservation and evolution of microsatellite loci in primate taxa. *Am J Primat* **50**: 205-214.

Colson, I. and Goldstein, D. B. (1999) Evidence for complex mutations at microsatellite loci in Drosophila. *Genetics* **152**: 617-627.

Day, J. C. and Ready, P. D. (1999) Relative abundance, isolation, and structure of phlebotomine microsatellites. *Insect Mol Biol* **8**: 575-580.

Engelke, D. R., Krikos, A., Bruck, M. E. and Ginsburg, D. (1990) Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. *Analytical Biochemistry* **191**: 396-400.

FitzSimmons, N. N., Moritz, C. and Moore, S. S. (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol Biol Evol* **12**: 432-440.

Fownes, S. L. (1999) Meadow suitability and quality for the alpine butterfly, *Parnassius smintheus*, in the east slopes of the Canadian Rockies. M.Sc Thesis. University of Alberta. 197 pp.

Glenn, T. C., Stephan, W., Dessauer, H. C. and Braun, M. J. (1996) Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. *Mol Biol Evol* **13**: 1151-1154.

Grimaldi, M. and Crouau-Roy, B. (1997) Microsatellite allelic homoplasy due to variable flanking sequences. *J Mol Evol* **44**: 336-340.

Grist, S. A., Firgaira, F. A. and Morley, A. A. (1993) Dinucleotide repeat polymorphisms isolated by the polymerase chain reaction. *BioTechniques* **15**: 304-309.

Harper, G. L., Piyapattanakorn, S., Goulson, D. and Maclean, N. (2000) Isolation of microsatellite markers from the Adonis blue butterfly (*Lysandra bellargus*). *Mol Ecol* **9**: 1919-1952.

Karhu, A., Dieterich, J. -H. and Savolainen, O. (2000) Rapid expansion of microsatellite sequences in pines. *Mol Biol Evol* **17**: 259-265.

Keyghobadi, N., Roland, J. and Strobeck, C. (1999) Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Mol Ecol* **8**: 1481-1495. **Chapter 2 of this volume.**

Kijas, J. M. H., Fowler, J. C. S., Garbett, C. A. and Thomas, M. R. (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques* **16**: 656-662.

Kostia, S., Varvio, S., Vakkari, P. and Pulkkinen, P. (1995) Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome* **38**: 1244-1248.

Li, W.-H. and Graur, D. (1991) Fundamentals of Molecular Evolution. Sinauer Associates, Inc. Sunderland, Massachusetts. 285 pp.

Ma, Z. Q., Röder, M. and Sorrells, M. E. (1996) Frequencies and sequence characteristics of di-, tri-, and tetra-nucleotide microsatellites in wheat. *Genome* **39**: 123-130.

Makova, K. D., Nekrutenko, A. and Baker R. J. (2000) Evolution of microsatellite alleles in four species of mice (Genus *Apodemus*). *J Mol Evol* **51**: 166-172.

Meglécz, E. and Solignac, M. (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas* **128**: 179-180.

Navajas, M. J., Thistlewood, H. M. A., Lagnel, J. and Hughes, C. (1998) Microsatellite sequences are under-represented in two mite genomes. *Insect Mol Biol* **7**: 249-256.

Nève, G. and Meglécz, E. (2000) Microsatellite frequencies in different taxa. *Trends Ecol Evol* **15**: 376-377.

Ortí, G., Pearse, D. E. and Avise, J. C. (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proc Natl Acad Sci USA* **94**: 10745-10749.

Paetkau, D. and Strobeck, C. (1994) Microsatellite analysis of genetic variation in black bear populations. *Mol Ecol* **3**: 489-495.

Paetkau, D. and Strobeck, C. (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol Ecol* **4**: 519-520.

Paetkau, D. (1999) Microsatellites obtained using strand extension: An enrichment protocol. *BioTechniques* **26**: 690-697.

Palo, J., Varvio, S., Hanski, I. and Vainola, R. (1995) Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. *Hereditas* **123**: 295-300.

Pfeiffer, A., Olivieri, A. M. and Morgante, M. (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* **40**: 411-419.

Primmer, C. R. and Ellegren, H. (1998) Patterns of molecular evolution in avian microsatellites. *Mol Biol Evol* **15**: 997-1008.

Raymond, M. and F. Rousset. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity* **86**: 248-249.

Reddy, K. D., Abraham, E. G. and Nagaraju, J. (1999) Microsatellites in the silkworm, *Bombyx mori*: abundance, polymorphism, and strain characterization. *Genome* **42**: 1057-1065.

Rico, C., Rico, I. and Hewitt, G. (1996) 470 million years of conservation of microsatellite loci among fish species. *Proc Roy Soc Lond B* **263**: 549-557.

Röder, M. S., Plaschke, J., König, S. U., Börner, A., Sorrells, M. E., Tanksley, S. D. and Ganal, M. W. (1995) Abundance, variability and chromosomal location of microsatellites in wheat. *Mol Gen Genetics* **246**: 327-333.

Rohwedel, J., Weichenhan, D., Meier, C. and Traut, W. (1993) Different modes of hypervariability in $(GATA)_n$ simple sequence repeat loci. *Insect Mol Biol* **2**: 49-58.

Schlötterer, C., Amos, B. and Tautz, D. (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**: 63-65.

Schlötterer, C. and Tautz, D. (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* **20**: 211-215.

Smith, D. N. and Devey, M. E. (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome* **37**: 977-983.

van Oppen, M. J. H., Rico, C., Turner, G. F. and Hewitt, G. M. (2000) Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. *Mol Biol Evol* **17**: 489-498.

Viard, F., Franck, P., Dubois, M., Estoup, A. and Jarne, P. (1998) Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *J Mol Evol* **47**: 42-51.

Weber, J. L. (1990) Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. *Genomics* **7**: 524-530.

Wolfe, K. H., Sharp, P. M. and Li, W.-H. (1989) Mutation rates differ among regions of the mammalian genome. *Nature* **337**: 283-285.

Zardoya, R., Vollmer, D. M., Craddock, C., Streelman, J. T., Karl, S. and Meyer, A. (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc Roy Soc Lond B* **263**: 1589-1598.

Table 6-1. Novel microsatellite loci useful for genotyping in *Parnassius smintheus*. FAM, HEX, and TET are fluorescent dye labels (Applied Biosystems). Data are for 557 individuals from 17 sampling sites. Size range of polymerase chain reaction (PCR) product is given in base pairs (bp). Null allele frequency was calculated using the EM algorithm. Expected heterozygosity (H_e) was calculated as $(1-\Sigma p_i^2)(2N/2N-1)$. Null allele frequency and H_e were averaged over sampling sites. * GTTT was added to the 5' end of the unlabelled primer of Ps163 to promote non-templated 3' adenine addition in PCR (Brownstein *et al.*, 1996). †For locus Ps165 all alleles greater than 250bp in length were pooled into the 250-bp size class, as they could not be sized unambiguously.

Locus name	Primer sequences (5' to 3')	Size Range	No. of non-null alleles	Mean frequency of null allele	Mean H_e
Ps76	**FAM** GGCAAATACCCTCCCTA GTAACGCTCAGTAAATCTGC	260 - 364	29	0.250	0.836
Ps163	**TET** CATTACCGAAACACGCACTT GTTTGCCAGGTCACGTTTAGGA*	283 - 316	9	0.044	0.282
Ps165	**HEX** CATGCGTAAATGTTGTAA CTAAACTAGGCGACGAAC	180 - >250†	36	0.358	0.825
Ps262	**TET** TTTGGTGTGTGCAAATGAAA TGCGACTGGATGGGATT	71 - 175	33	0.032	0.916

Table 6-2. Amplification results for all microsatellite loci isolated from *Parnassius smintheus* for which polymerase chain reaction (PCR) primers have been designed. Standard isolation protocol refers to an unenriched protocol involving bacterial transformation and plaque screening (Paetkau and Strobeck, 1994). PCR-based protocol refers to the method of Grist *et al.* (1993). 1 = isolation described in Keyghobadi *et al.* (1999). 2 = isolation described here. 'Usable' loci amplified cleanly with only one or two bands per individual, displayed allelic variation, and yielded products in most individuals. * Locus Ps162 was originally deemed useful for genotyping populations in Kananaskis, Canada (Keyghobadi *et al.* 1999) where it had null alleles; it was later found to display up to three bands per individual in some populations from Banff National Park, Canada. Mg^{++} refers to magnesium ion concentration in the PCR cocktail. T_a refers to PCR annealing temperature.

Locus name	Isolation protocol	Microsatellite repeat region	Usable?	Amplification pattern	PCR conditions tested		No. of primers tested		Accession No.
					Mg^{++} (mM)	T_a (°C)	Strand 1	Strand 2	
Ps44	standard[1]	$(TG)_5(TA)_3$	No	Multiple bands	0.5-2.0	54-60	1	1	AY048083
Ps50	standard[1]	$(GATA)_{10}$	Yes	Null alleles	2.0	54	1	1	AF133661
Ps58	standard[1]	$(GCGT)_{12}(GT)_8$	No	Multiple bands	0.5-2.0	54-60	2	2	AY048084
Ps67	standard[1]	$(GT)_{34}$	No	Multiple bands	0.5-2.0	54-60	2	2	AY048085
Ps92	standard[1]	$(ATT)_7...(GT)_6$	No	Invariant	2.0	54	1	1	AY048086
Ps106	standard[1]	$(GT)_6G_{10}$	No	Multiple bands	0.5-2.0	54-60	1	1	AY048087
Ps81	PCR-based[1]	$(GT)_{10}$	Yes	Null alleles	0.5-5.0	48-60	3	2	AF133662
Ps85	PCR-based[1]	$(GT)_{10}$	Yes	Null alleles	1.0-7.0	47-56	3	2	AF133663
Ps135	PCR-based[1]	$(TG)_7$	No	Null alleles	0.5-5.0	48-60	3	2	AY048088
Ps162	PCR-based[1]	$(GT)_6$	*	Null alleles/Multiple bands	1.0-2.5	54-62	1	1	AF133664
Ps76	PCR-based[2]	$(TTG)_3(TG)_8$	Yes	Null alleles	2.0	54	2	1	AY048089
Ps112	PCR-based[2]	$(TG)_6A(TG)_3$	No	No product of expected size	2.0	54	2	1	AY048090
Ps163	PCR-based[2]	$(GT)_4$	Yes	Null alleles	2.0	54	1	1	AY048091
Ps165	PCR-based[2]	$(GT)_{12}$	Yes	Null alleles	1.0-4.0	54-58	2	1	AY048092
Ps180	PCR-based[2]	$(TG)_{13}(AG)_{14}$	No	Multiple bands	1.5-2.0	54	1	1	AY048093
Ps206	PCR-based[2]	$(TG)_{11}$	No	Multiple bands	1.0-2.0	54-63	2	2	AY048094
Ps248	PCR-based[2]	$(GCGT)_4...(GCGT)_7$	No	No product of expected size	2.0	54	1	1	AY048095
Ps262	PCR-based[2]	$(TG)_{13}$	Yes	Null alleles	2.0	54	1	1	AY048096

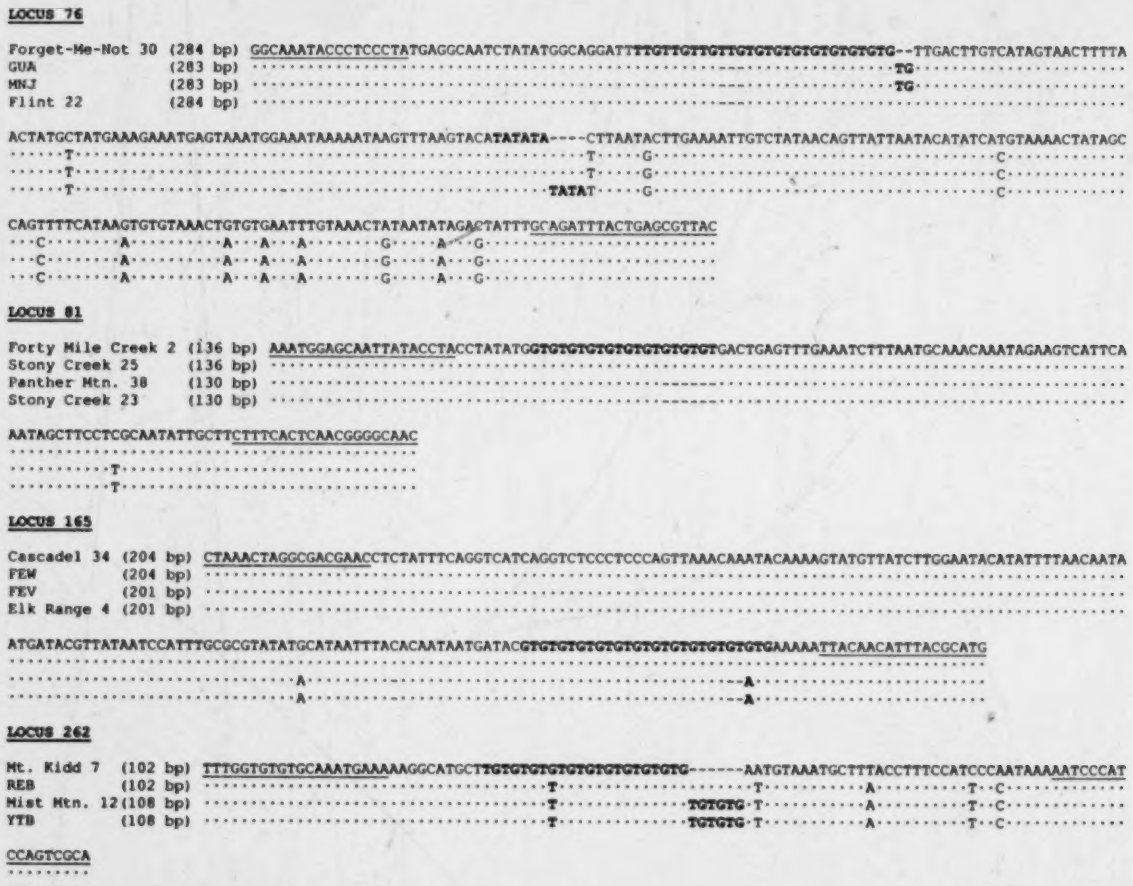


Figure 6-1. Aligned sequences of four alleles representing two size classes for each of four microsatellite loci from *Parnassius smintheus*. Underlined sequences are primers for amplification from *P. smintheus*. Dots indicate identical sequence. Dashes indicate indels. Repetitive regions are in bold. Accession numbers for sequences are, locus 76: AY048105-AY048108, locus 81: AY048109-AY048112, locus 165: AY048097-AY048100, locus 262: AY048101-AY048104.